Sensors and Materials, Vol. 37, No. 4 (2025) 1311–1331
MYU Tokyo

S & M 3989

Realizing Short-term Disease Forecasting in Crops via Multimodal Monitoring with Leaf-underside-sensing Agricultural Robot

Kenji Terada,[1,2*] Shigeyoshi Ohno,[1] and Kaori Fujinami[2**]

[1]Polytechnic University of Japan, 2-32-1 Ogawa-nishimachi, Kodaira-shi, Tokyo 187-0035, Japan
[2]Department of Bio-Functions and Systems Science, Tokyo University of Agriculture and Technology, 2-24-16 Nakacho, Koganei City, Tokyo 184-8588, Japan

(Received October 28, 2024; accepted January 28, 2025)

Keywords: smart agriculture, Internet of Things, artificial intelligence, precision agriculture, agricultural robot

This study was focused on forecasting diseases in fruit trees five days in advance using supervised machine learning. This involved photographing the undersides of leaves and sensing environmental conditions using agricultural robots. A leaf underside disease classifier that achieved 0.90 accuracy and 0.91 recall based on 330 images collected by the robot-mounted camera was developed. The classifier's results were utilized for binary classifications to predict disease occurrences. This innovative approach aims to enhance disease management in agriculture. Using objective variables in the leaf underside disease classification and feature-increase method, we analyzed disease forecasting methods through the comparison of machine learning models, sensor types, and dataset durations required for training the models. As a result, we clarified the changes in the accuracy of the predicted number of days for each machine learning model. The recall when using the dataset collected by the robot over 16 days was 0.980. Furthermore, we confirmed that the characteristics unique to each farm appeared in the forecast for each sensor used in the observations.

1. Introduction

"Zero hunger" is a theme long pursued by humankind. The world's agricultural markets are facing an unprecedented food production crisis owing to rapid changes in weather conditions and soaring prices of agricultural chemicals and fertilizers.(1) The rapid decline in the number of farmers, particularly in developing countries, the lack of successors, and labor force shortages have made it an urgent issue to increase the average size of arable land per farmer. Significant attention has been paid to smart agriculture and precision agriculture,(2) which aim to reduce the number of farmers by using data. It will become increasingly important for farmers to use the Internet of Things (IoT), artificial intelligence (AI), and robotic technology, represented by drones, to visualize information on farms and respond to each crop individually.(3) Most farmers

*Corresponding author: e-mail: k-terada@uitec.ac.jp
**Corresponding author: e-mail: fujinami@cc.tuat.ac.jp
https://doi.org/10.18494/SAM5402

ISSN 0914-4935
https://myukk.org/

generally apply pesticides to an entire farm at predetermined times using a calendar spraying system. A low-cost spot spraying system used to detect infested sites in greenhouses and orchards can be realized for pesticides if farmers can precisely identify diseases at the crop level within a farm. In addition, typical crop pests, such as beetles, spider mites, and lepidopteran pests, parasitize and feed on the underside of leaves.[4] The analysis of leaf health is important for healthy plant growth.[5] The underside of leaves is an important target for analysis because insects, which are the main cause of disease, often congregate on the underside of leaves.

Identifying region-specific diseases requires specialized knowledge and expertise acquired through rigorous agronomic research; it is often referred to as tacit knowledge. This underscores the importance of timely and effective disease control measures. Failure to address plant viral diseases in one area can lead to unchecked proliferation, potentially causing widespread secondary damage to other crops on the same farm. Pests account for 40% of all food loss.[6] Disease control relies mainly on the farmer's accumulated experience, which is not always stable and requires a long time and experience to maintain a high degree of accuracy.[7] Most farmers spray pesticides and herbicides over the entire crop for pest control. However, crops are exposed to excessive amounts of chemicals that may adversely affect human health.[8] Therefore, the detection of crop-specific diseases is required. If disease detection can be performed on a crop-by-crop basis, the spot application of pesticides will become possible. Image-based disease detection is effective for disease forecasting. However, the main cause of crop disease is viral infection. Just as the common human cold occurs after several days of physical illness and viral infections caused by temperature and other weather changes, visually visible plant diseases are often internally infectious to crops a few days before the disease is detected. Therefore, systems that can monitor the disease and growth status of each crop individually as much as possible and conduct spot treatments such as pesticide sprays and vibration are required. In smart agriculture, fields can be observed by means of human senses and memory, fixed sensors for robust observation that are not easily affected by weather conditions, and drones for wide-area observation above the field. In this study, we aimed to use a sensing agricultural robot that can travel over uneven terrain in the field to assist farmers with various observations and analyses, such as imaging the undersides of crop leaves and measuring environmental values on the ground a few meters above detailed surface mapping, which are difficult to accomplish manually or using fixed sensors or aerial drone observation methods. Disease symptoms on the underside of leaves have been recognized for centuries.[9] Although it is widely acknowledged that observing the undersides of leaves near unstable soil environments is crucial, there have been concerns regarding delayed detection owing to the difficulty of visual inspection by humans.

In this study, we focused on the capabilities of the sensors and cameras mounted on agricultural robots. In general, the following processes of crop disease development are of interest.[10,11] Plants, similar to humans, develop diseases after being exposed to a stressful environment for several days. Diseases are caused by sudden environmental changes that persist for several days. The farmers' request is to achieve a forecast of crop disease for the next two days with a probability of 50% or more. The novelty or significance of this study is that it predicts the symptoms of diseases that appear on leaves before they can be seen. The mechanisms underlying disease development are complex and vary greatly depending on the surrounding

environment, which varies from farm to farm. This study is aimed at detecting diseases at an early stage using machine learning (ML) by focusing on environmental changes that trigger disease emergence. By the time a person notices the disease, the source has most likely spread throughout the field, and it is too late to take countermeasures once symptoms appear on the branches and leaves. Therefore, we identify techniques to forecast disease before early symptoms appear on branches and leaves, using multimodal information that comprises measurements of environmental values by ground rovers and the analysis of images of the undersides of leaves. The outcomes of our study will support farmers' decision-making processes. The contributions of this study are as follows.

(1) We suggest an integrated forecast method using information with different granularities such as the wind direction of a farm region from a government agency, the wind speed from a sensor fixed in a farm, and the environmental values collected by a ground sensing agricultural robot to enable disease forecasting per crop.
(2) We propose an automatic annotation method for objective variables in the dataset to train the forecast model, which is created by a disease classifier using camera images of the underside of leaves.
(3) We analyze the impact of parameters that affect the performance of forecasts and show that the disease forecast of 5 days ahead based on 16 days of observations has the best recall of 0.980.

This paper is organized as follows: In Sect. 2, we discuss related studies. In Sect. 3, we first provide a comprehensive overview of the issues identified in the related research. We propose a sensing agricultural robot capable of multimodal observation to address these issues. and we establish a feature-increase method and the handling of unbalanced data for forecasting. In Sect. 4, we present the experimental environment. We describe the evaluation method used for verification, the conditions under which the training data were collected, and the annotation of leaf underside images for disease classifier training. In Sect. 5, we present an analysis of evaluation items and evaluation methods through comparisons of ML models, sensor types, and dataset durations required for training the models, as well as the classification of diseases on the basis of the state of the leaf underside. Finally, in Sect. 6, the conclusions and future direction are given.

2. Related Works

2.1 AI-based farm management system

In recent years, research on agricultural support using sensing technology, focusing on field observations, has produced useful results using AI. The Japanese Ministry of Agriculture, Forestry, and Fisheries (MAFF) offers a public cloud service called WAGRI, which provides data and programs useful for agriculture, such as data for forecasting weather, farmland, and yield.[12] WAGRI provides various open and big data, such as breeding forecasts and pest diagnosis via Web APIs, and implements services to increase productivity in agriculture. Wakabayashi[13] advocated agricultural ICT services that employed AI and IoT technologies.

They researched the use of AI to evaluate tea leaves by image analysis so that harvesting could be conducted efficiently at the appropriate time. By establishing the standards, various local governments, universities, and companies collaborated to consolidate the data. The main issue was the lack of a pest database. As of 2020, the database of pests and diseases is insufficient for pest forecasting. Mori[14] introduced Plantect, a smart agricultural solution for greenhouse cultivation that used AI to predict the risk of disease infection. With the cooperation of testing and research institutes and growers across the country, data were collected in more than 10000 greenhouses and examined to accumulate information on temperature, humidity, disease name, date of disease onset, chemical application records, and field and cultivation information related to disease outbreaks. The predicted temperature and humidity in the greenhouses up to two days after the outbreak were applied to a disease infection risk forecast model using random forest (RF) as an ML method, and the infection risk up to two days after the outbreak was predicted. Fenu and Malloci[15] used historical weather data provided by the Agenzia Regionale per la Protezione Ambientale (ARPAS) agency from the Cagliari location over three years (2016–2018) to examine whether weather data from regional weather stations could be used to predict the risk indicators for potato late blight. Artificial neural network (ANN) and support vector machine (SVM) classification achieved 96 and 98% forecast accuracies, respectively. Gianni and Maridina *et al.*[16] performed an analysis and classification of forecasting models for crop diseases over the past 10 years (2010–2020). They pointed out that, in terms of the technical level in this field, only a limited number of contributions have been published from 2010 to the present day. They also pointed out that the majority of the literature focuses on a small number of pathogens and crops, and that of these, only a few consider data from various heterogeneous sources to predict disease outbreaks.

We believe that the AI-based forecasting challenge lies in the difficulty and precision of data collection. The collection of images is performed by farmers using smartphones or other means, which requires a great deal of labor. The detailed management of observation conditions, such as temperature and humidity, at the time of disease imaging during image collection is insufficient. In addition, forecasts using fixed sensors and weather information are made on a farm-by-farm basis, and precise forecasts for each crop or 1-m-square area have not been implemented. Therefore, for more precise observations "per crop," it would be possible to provide more useful information for agriculture if environmental conditions, such as ambient temperature, humidity, and carbon dioxide, could be automatically measured at the time of image observation and if multimodal observations could be achieved for easier disease forecasting.

Plant-image-based disease identification applications exist in several markets, particularly smartphones.[17] Qadri *et al.*[18] proposed plant disease detection and segmentation using end-to-end YOLOv8. This research intends to present a deep learning solution to the detection and segmentation of plant leaf diseases. Leveraging the PlantVillage and PlantDoc datasets to train the Ultralytics YOLOv8 model from end to end enables complex feature extraction from images, leading to the precise identification of plant leaf diseases. The evaluation results for the YOLOv8 approach are validated by prominent statistical metrics like precision, recall, mAP50 and mAP50-95 values, and $F1$-score, which resulted in 99.8, 99.3, 99.5, 96.5, and 0.999 for the bounding box and 99.1, 99.3, 99.3, 98.5, and 0.992 for the segmentation mask, respectively. Lou

et al.[19] successfully demonstrated the effectiveness of YOLOv5 as an image classifier for detecting larvae, moths, and grasshoppers in agricultural fields by changing the CIoUloss function of YOLOv5 to the EIoUloss function to improve the accuracy. Shi *et al.*[20] observed not only that several agricultural pest identification software programs are available on the market but also that their identification accuracies were unstable in different fields. Three hundred photographs of each tomato disease were randomly selected to solve the dataset bias with various numbers of training data depending on the disease, on the basis of EfficientNetV2 transfer learning. Pictures were randomly selected and subjected to mirror-image transformations in the horizontal and vertical directions. Nine common tomato diseases were correctly identified, with mAP values reaching 0.98, indicating that learning was required in different fields. Disease classification using YOLO and EfficientNet demonstrated the effectiveness of the classifiers that could be applied to farming. However, related studies have shown that it is necessary to use images of the same species; most of the images used for training in these studies were leaf surface images. It is difficult to determine whether these datasets will be effective, even if they are used in this study with transfer learning.

2.2 Unmanned aerial vehicle (UAV)

The advent of drones, also known as UAVs, is promoting more efficient pesticide and fertilizer applications, as well as farm sensing to monitor growth conditions.[21] Chenghai[22] successfully detected and mapped numerous crop diseases using drones and satellite images. Noting the need to use advanced image sensors and analytical methods to distinguish diseases from other confounding factors in a wide range of images, they stated that early detection was challenging. Sugiura[23] visualized variations in growing conditions by acquiring normalized vegetation indices using a multispectral camera from the air over a large farm in Hokkaido, Japan, using an aerial drone. He highlighted the necessity of aligning the acquired images with ground-truth datasets as a challenge and noted that, based on the type of disease, it is often difficult to obtain the level of detection accuracy equivalent to a visual evaluation on the ground, particularly with a skilled eye, and that a universal solution cannot be expected.

UAVs are limited in that they cannot fly in rain, and strong winds, at night, or in urban areas. Although it is simple to imagine the effectiveness of nighttime observations in detecting nocturnal pests, such as weevils, it is difficult for UAVs to fly at night, as laws and regulations in several countries are becoming stricter every year. We believe that ground-based observations will be required to achieve precision agriculture.

2.3 Unmanned ground vehicle (UGV)

In a study focused on the importance of visual evaluation on the ground, a ground rover called an UGV is utilized to drive over the ground (Table 1).[24] Grimstad *et al.*[25] considered the diversification of farm environments a challenging issue. They employed a four-wheel-drive robot with a track width of 1.5 m and a mass of less than 200 kg and also equipped with a passive suspension to ensure drivability on rough terrain. This enabled the monitoring and UVB

Table 1
UGV characteristics.

Researcher	Robot	Feature	Target crops
Lars *et al.*	Thorvald II	monitoring and UVB treatment	cucumber, tomato
Gonzalez-De-Santos *et al.*	unmanned vehicle	four-wheeled with legs, RTK-GNSS	corn
Quaglia *et al.*	Agri.q	mapping, monitoring, etc.	unspecified

treatment of cucumber and tomato plantations in Norway and the UK. Gonzalez-De-Santos *et al.*[26] proposed an unmanned vehicle for smart farms and designed a four-wheeled rover with legs as a steering mechanism. However, the difficulty in motion control is a drawback. LeVoir *et al.*[27] achieved the same accuracy as RTK-GNSS, and Quaglia *et al.*[28] developed Agri.q, an innovative and sustainable robot for precision agriculture. Equipped with functions such as mapping, monitoring, and sample collection, it navigated unstructured environments through its unique design and powerful sensors. Weight-optimized mechanics can enhance sustainable operations. In this study, we give details of the features of Agri.q and its energetic performance under various conditions.

The issue with agricultural rovers outside Japan is that many are large scale. To traverse uneven and rough terrains, many parts, such as dampers and steering mechanisms, are required, making larger rovers inevitable. The smaller the rover, the more it is affected by uneven terrain, making it difficult to obtain stable observations such as camera images. However, the size of a passenger car would be difficult to apply in mountainous or residential areas, such as Japan, which has many densely populated cultivation areas.

3. Disease Forecast Method and Apparatus

3.1 Model overview

According to a 2024 report by Tammina *et al.*,[29] the following are considered important in AI disease prediction: “robots”, “disease images”, “sensing”, “imbalanced disease data”, and “applicable algorithms”. Autonomous agricultural equipment using robots, combined with AI, is said to provide effective and accurate real-time solutions for disease diagnosis and management. Disease images are used to record the symptoms and signs of disease. Such images cover a variety of environments, lighting types, and plant species, ensuring the durability and flexibility of the model. Information from environmental sensors such as relative humidity and temperature is used to improve the disease prediction model. The accuracy of the prediction is improved by relating these environmental factors to the incidence of disease. Imbalanced data can affect the ability to identify diseases. The balance of disease classification can be achieved by over-sampling minority classes such as diseases or under-sampling dominant classes. Information on past disease outbreaks, weather trends, and agricultural practices is useful for building disease prediction algorithms. Using patterns and trends obtained from past data, it is possible to predict disease risk. These elements are essential for developing robust AI-driven disease prediction systems, ultimately enabling proactive and precise disease management in agriculture.

On the basis of the above elements in disease prediction, we designed a data processing pipeline shown in Fig. 1. Using a sensing agricultural robot (Sect. 3.2), images of leaf undersides and environmental data are collected for constructing a disease forecast model and operating the forecast. Training data are generated by combining the feature values using each observation value and its neighborhood values with disease classification results of a few days ago (Sect. 3.3) while addressing imbalanced data (Sect. 3.4).

The environmental values such as temperature, humidity, CO_2, and UV within the farm do not have a uniform distribution because the growth conditions of each crop differ. Figure 2 shows a heatmap of the humidity and CO_2 observed by the sensing agricultural robot at the mulberry farm used in this study. The air did not reach the center of the farm owing to the direction of the wind, which varied the measurement values. By integrating accurate information from the various sensors installed in the sensing agricultural robots, we aim to capture the surrounding environment of each crop by capturing even the slightest differences in environmental factors and forecast disease.

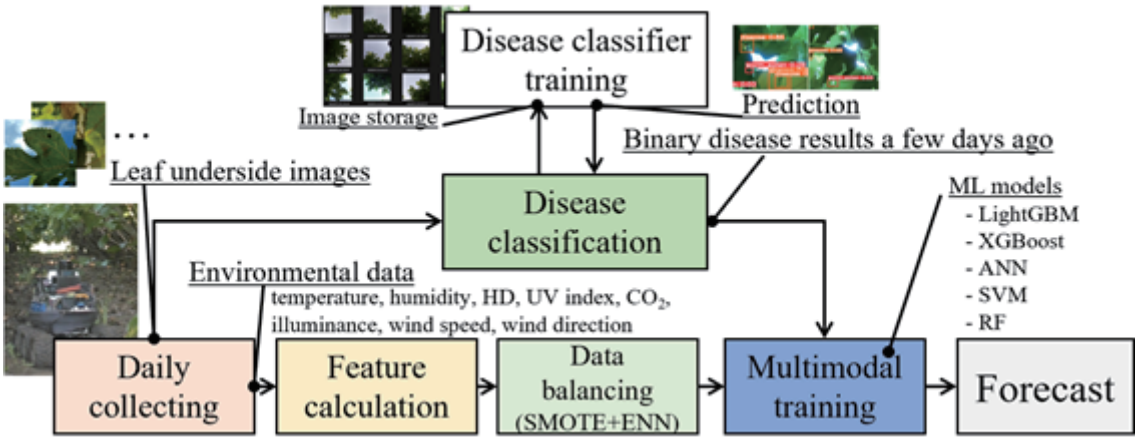


Fig. 1. (Color online) Data processing pipeline in the disease forecast.

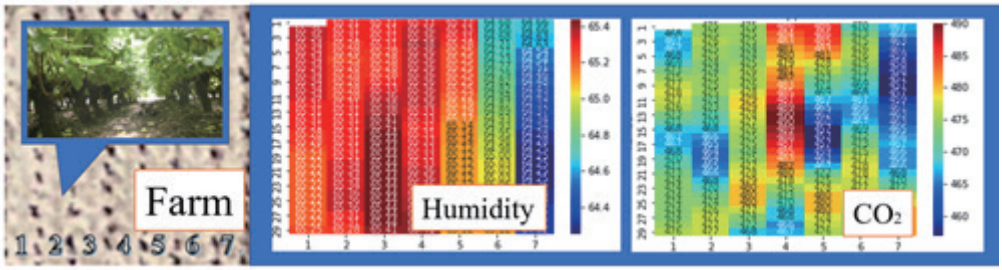


Fig. 2. (Color online) Heat map based on the observation results of the sensing agricultural robot (average values for each observation point from July 7 to 12, 2022).

3.2 Equipment

Multimodal disease observation methods aim to forecast disease outbreaks before the appearance of disease symptoms in crops. A robot running on the ground can operate effectively below the underside of the leaves. However, several crops on the farm were planted at regular intervals. The challenge is to achieve constant-speed performance on uneven terrain for crop monitoring. Therefore, in this study, we used a sensing agricultural robot with an electric six-wheel drive-related suspension structure (Fig. 3, Table 2).[30]

Table 3 shows the environmental sensors, effect on crop, and product names for the sensing agricultural robots used in this study. Next, the position of each plant (horizontal: row; vertical: column) recorded at regular intervals while the robot was traveling at a constant speed around the farm was used as an exogenous data source. Furthermore, we used the wind direction (WD) at observation points published by the Japan Meteorological Agency because wind is a source of airborne contagion owing to viral dispersal. In addition, a camera was equipped to observe the undersides of leaves where disease-causing insects gather, and ground truth labels for training the forecast model were generated from the collected images. The robot's observation unit uses a Raspberry Pi 4B+ with the Linux-based Raspbian OS installed. Each sensor is connected to the Raspberry Pi 4B+.

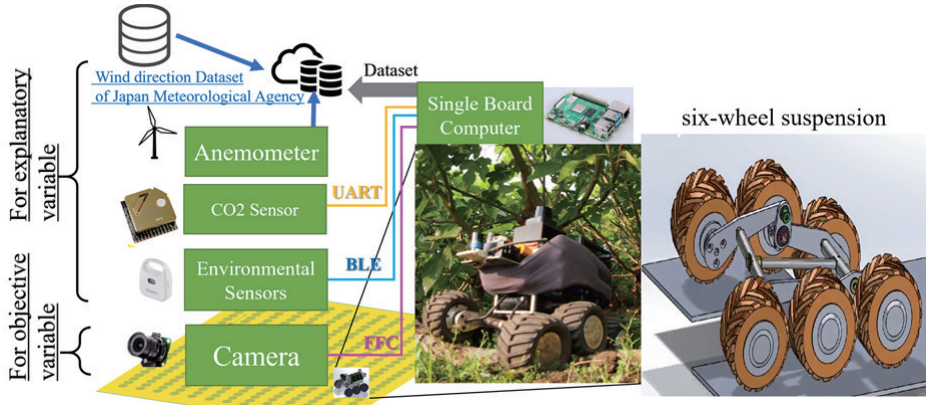


Fig. 3. (Color online) Disease forecasting system with sensing agricultural robot.[30]

Table 2
Specifications of sensing agricultural robot.

Width (mm)	294
Depth (mm)	350
Height (mm)	310
Weight (g)	7400
Wheel diameter (mm)	65
Maximum capacity (kW)	3
Maximum speed (km/h)	30.3
Battery life (min)	63.3

Table 3
Sensors for farm observation

Sensor	Effect on crop	Product name
Temperature	The standards for farmers to predict the occurrence and spread of diseases	2JCIE-BL (OMRON Corporation)
Humidity	The excessive humidity affects the formation of spores in pathogens and their invasion of plants.	
Humidity deficit (HD)	The water vapor absorptivity of plants changes with temperature.	
Ultraviolet (UV)	Harmful light to plants and causes disease	
Illuminance	The intensity of sunlight that can be received by a person nearby, and sunlight has a direct effect on plant growth and development.	
CO_2	The indicator of plant activity	S-300 (ELT SENSOR Corp.)
Wind Speed (WS)	A source of airborne contagion owing to viral dispersal	CHE-WD2 (SANWA SUPPLY)

3.3 Formulation of disease forecast model

In this study, the forecast model is trained using data collected from several days before the prediction, rather than using a large amount of data collected over a long period.

A recent study that is similar in purpose to this research is the time series prediction using long short term memory (LSTM) by Honda *et al.*[31] This study used a dataset of temperature and humidity data observed every minute, and the sliding window method was used to obtain the estimated and predicted values for each time series data by shifting the length of the time series data by 1 min, with the model accuracy verification set to 5 min. In this study, we applied LSTM to the time series data collection obtained from the farmer described in Sect. 4.2, and conducted a forecast of the next two days using 6300 dimensions of five days. The recall was 0.000, which indicates that no diseased areas were found. LSTM is excellent at capturing long-term dependencies that span several years; however, it is computationally expensive, and complex models are difficult to perform without an outrageous amount of training data.

Therefore, we defined the disease forecast model that can be applied to simple, supervised ML models that are likely to perform well even with a small amount of training data. The forecast is formulated by Eq. (1), in which $I_{s,\Delta t}(\overrightarrow{F_t}, d)$ and $y_{t+\Delta t}$ represent a forecast model trained using the data from s days to forecast the state of Δt days ahead, input features for day t calculated from d days as a vector, and a predicted state of disease after Δt days from day t. The result takes a value of 1 or 0 depending on whether disease is predicted to occur or not. Thus, the forecasting task is a binary classification. The model is constructed by supervised ML using s-pairs of labels and features (Fig. 4 right). A lag of Δt is introduced between the time of acquisition of the pair of labels and features because this is a forecast. The labeling method is presented in Sect. 4.3.

$$y_{t+\Delta t} = I_{s,\Delta t}(\overrightarrow{F_t}, d) \tag{1}$$

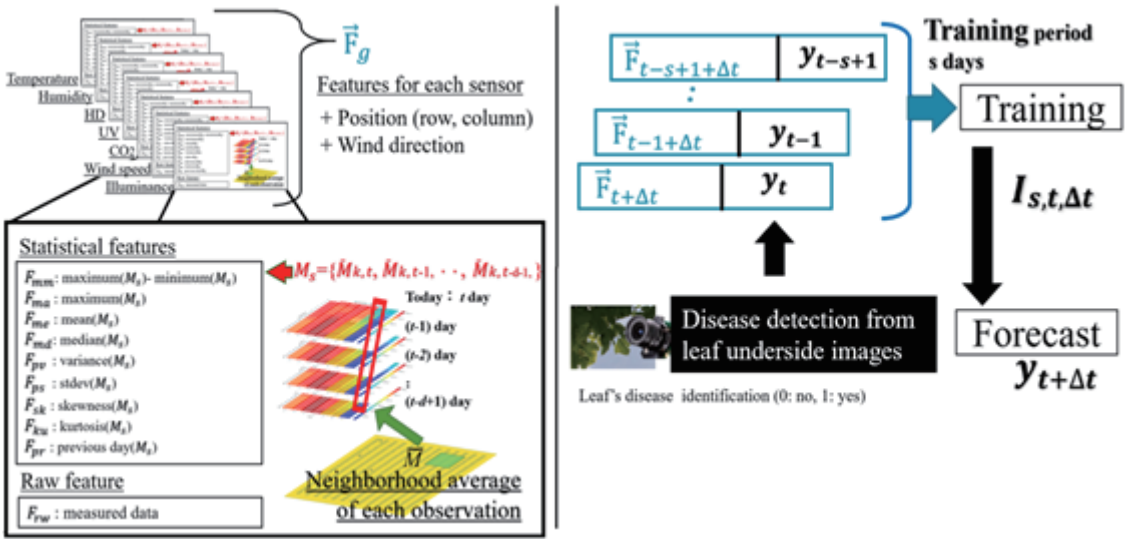


Fig. 4. (Color online) Feature generation method.

Figure 4 (left) shows the feature calculation method. $\overrightarrow{F_g}$ represents a set of features from sensor $g \in$ {temperature, humidity, HD, UV index, CO_2, wind speed, illuminance}, which consists of statistical features and the measured raw data (F_{rw}). Nine statistical features were introduced: the difference between the maximum and minimum values (F_{mm}), maximum value (F_{ma}), mean (F_{me}), median (F_{md}), variance (F_{pv}), standard deviation (F_{sd}), skewness (F_{sk}), kurtosis (F_{ku}), and difference in value from the previous day (F_{pr}). Each statistical feature was obtained from a set of data ($\{\bar{M}_{k,t}\}$) of d days, including the day of observation t for each position k. Each $\bar{M}_{k,t}$ value was obtained by averaging the measurements from position k and its eight neighboring positions. Thus, ten features, nine from statistical operations and one from raw measurements, were obtained for each of the seven sensors. In addition, the position of measurement (k), that is, the row and column in the farm coordinates, and the wind direction were included in the feature vector. In total, a 73-dimensional vector was provided to the forecast model to obtain the prediction of Δt days ahead.

3.4 Handling of unbalanced data

The number of disease-forecast sites was generally smaller than that of undetected sites. Thus, the observed data were considered to be unbalanced, as shown in Fig. 5. Therefore, we leveraged a data balancing technique, the synthetic minority oversampling technique with edited nearest neighbors (SMOTE + ENN),[32] which was a combination of oversampling and undersampling rigging techniques. Undersampling is a method used to reduce the amount of data of many classes. Oversampling is a method used to increase the amount of data of a small number of classes. Sample weight is a method that reduces the weight of samples close to the decision boundary. SMOTE increases the number of data points in a minority class by generating new data points using nearby data points. ENN uses the k-nearest neighbor method to examine

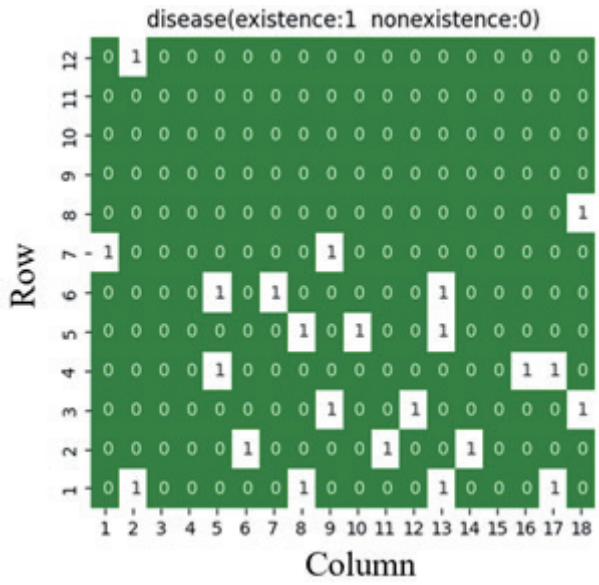


Fig. 5. (Color online) Example of a disease heatmap of an observed orchard.

the data points in the neighborhood of each data point; if the labels of the data points in the neighborhood do not match, the data point is deleted.

4. Experiment

4.1 Evaluation items and method

Recall represents the completeness of the model, which indicates how well the model identifies all true positives. In the evaluation method for ML in this study, recall, which can be calculated using Eq. (2), was applied. The precision was calculated using Eq. (3). True positive (*TP*) is when the model correctly predicts a positive outcome (e.g., identified diseased leaf). False positive *(FP*) is when the model incorrectly predicts a positive outcome (e.g., mistook a healthy leaf to be a diseased leaf). False negative (*FN*) is when the model incorrectly predicts a negative outcome (e.g., failed to identify a diseased leaf).

$$Recall = TP / (TP + FN) \tag{2}$$

$$Precision = TP / (TP + FP) \tag{3}$$

The reason for adopting recall is that the farmers who used this study were considered. Disease detection in farming is a high priority. For example, in the case of pesticide spraying, farmers want to avoid undetected disease owing to failure to detect the disease rather than spraying pesticides on areas where no disease exists. A high recall indicates that the model is less likely to miss diseased areas, which is crucial for farmers, as they prioritize detecting actual diseases over avoiding unnecessary pesticide use.

4.2 Training data collection

The experimental data required for this study were collected from an outdoor mulberry field. The mulberry trees were planted in a straight line at specific intervals, and the robot traveled through 12 rows of trees to make observations using the sensors and cameras described in Sect. 3.1. The robot had to be turned around manually because there was no space available for autonomous turning owing to the public road at the corner of the farm. By patrolling the farm, observations were made within 1 h after dawn when the stomatal opening was stable under clear skies. The robot must perform one daily patrol to reduce rutting in the soil. In addition, no measurements were taken when rain occurred because rain damages the soil. The speed was adjusted such that each row could be covered in 1 min, and 18 observations were made per tree row, excluding data from both ends. The total measurement time for the entire farm was approximately 12 min, as the robot traveled at a speed of 1 min per row and observed 12 rows. The farmland was uneven terrain with bumps created by small branches and animal carcasses. Data were collected in 2022 from different mulberry fields with the size of approximately 10 × 7 m^2. The detailed spatial scale within the farm is 70 m^3, with the height of a tree being 1 m. We collected data from days of observations on 25 days when it was not raining, from August 4 to September 25, 2022.

4.3 Annotation of leaf underside images for disease classifier training

We collected 330 pieces of information on chlorosis from 2592 observation points on farms adjacent to the fields where we collected datasets in the city from July 2 to July 22, 2022. The ground truth for the continuous training of the disease forecast model was determined by a disease classifier that used the images captured by the camera on the sensing robot. The disease classifier was constructed with YOLOv5 using underside leaf images, in which the leaf undersides with visible discoloration were designated as “diseased”.

The classifier constructed in YOLOv5 categorized the state as “disease” or “worm eaten” (Fig. 6). In this study, only the classification of discoloration was used and “disease” and others were ignored. Observations were conducted in the shade, so that the location of the disease was difficult to determine clearly. Therefore, we defined a disease detection score of 0.5 or higher as disease to expand the possibility of disease detection. Furthermore, in cases in which multiple diseases were detected in the captured images, the value with the highest score was used for ML. The highest classification score was used if more than one disease occurred on the leaf underside in the image. The classification score of the classifier was set to 0.5 as the threshold value, and disease information was added to the sensor observation points as the ground truth, with the value of one for disease detection and zero for no disease detection, without relying on human visual inspection. In this study, 330 underside leaf images were collected in an orchard from July 2 to 22, 2022, and rectangular annotation was applied to train the disease classifier. Annotation was performed using visual classification with square selection. The number of learning epochs was increased to improve accuracy. Finally, no improvement in accuracy was observed, even when the number of epochs exceeded 200. The accuracy and recall were 0.90 and 0.91,

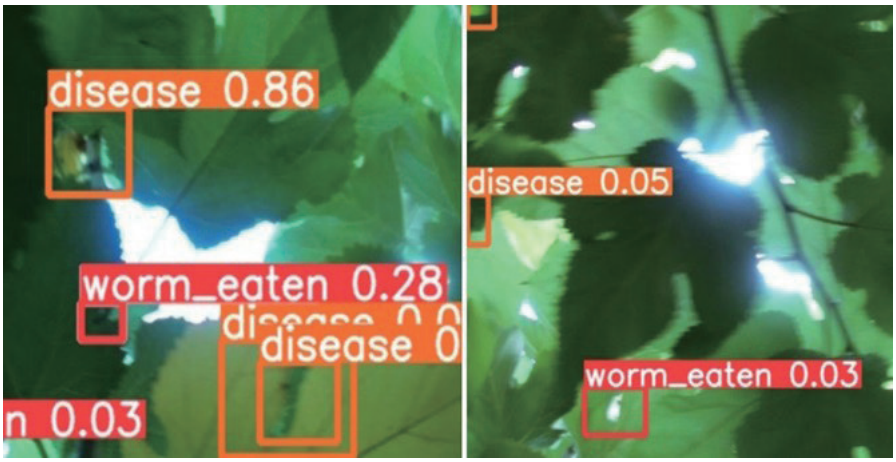


Fig. 6. (Color online) Disease classification by YOLOv5.

respectively. When the number of epochs was increased to 300, each score decreased by 0.002, so the number of epochs used in the automatic classifier was set to 200. The main reason for the high detection accuracy, even with a small number of images, was that the robot looked up when collecting images of the undersides of the leaves. The background of each observation point was a monochromatic image of the sky above, and the observation time was at dawn. Therefore, there was little effect from backlighting, resulting in few patterns in the detected images.

5. Results and Discussion

5.1 ML algorithms and unbalance data countermeasures

We used data from 17 days of observations on days when it was not raining, from August 4 to August 27, 2022. We assumed that there was one observation per day. The observation time was within 1 h of dawn, when the stomata open stably. There were a total of 263160 observation points. We compared the recall of each ML algorithm. To select a forecasting method, only the Δt value of zero in Eq. (1) was verified. The three methods compared were the index based on the observed environmental values alone, the index when SMOTE + ENN was applied, and the index when SMOTE + ENN plus the feature-increase method proposed in this study was applied. We compared the ML algorithms XGBoost,[33] LightGBM,[34,35] ANN, SVM,[36] and RF.[37] The average recall values are listed in Table 4. The average precision rates are listed in Table 5. If we focus only on the results in Table 4, we might assume that SVM with SMOTE + ENN is applied to the observed values because it has the highest average recall. However, the precision for the same item in Table 5 was 0.000. This is because the classifier incorrectly predicted that the disease was present at all observation points. However, this cannot be considered a precise disease forecast. Therefore, LightGBM with SMOTE + ENN and a statistical feature increase, which had the highest recall, was used, even though the precision was not 0.000. LightGBM with SMOTE + ENN exhibited the highest recall. One of the possible

Table 4
Average recall of each analysis algorithm.

Feature pattern	XGBoost	LightGBM	SVM	ANN	RF
1. Environmental observation values only	0.013	0.021	0.000	0.001	0.017
2. SMOTE + ENN added to 1	0.533	0.604	0.993	0.908	0.903
3. Statistical features added to 2	0.332	0.819	0.805	0.471	0.485

Table 5
Average precision of each analysis algorithm.

Feature pattern	XGBoost	LightGBM	SVM	ANN	RF
1. Environmental observation values only	0.001	0.001	0.000	0.001	0.011
2. SMOTE + ENN added to 1	0.034	0.028	0.000	0.008	0.000
3. Statistical features added to 2	0.095	0.213	0.088	0.081	0.095

reasons for the highest recall of LightGBM is that it forms decision trees leafwise, whereas LightGBM forms decision trees levelwise. LightGBM is a levelwise decision tree that tends to increase its complexity. Higher complexity indicates that the classification is progressing efficiently. However, there is a possibility of overfitting, in which too many adaptations are made to the dataset, thereby reducing its generality.

5.2 Disease forecast by each ML algorithm

Farmers from the companies that provided the sensing agricultural robots indicated that the forecast results within two days of the date of observation were effective for agricultural decisions. We compared the recall of supervised ML algorithms such as XGBoost, LightGBM, ANN, SVM, and RF by varying the numbers of training days s and lag days Δt. None of the ML algorithms sets hyperparameters to adjust at the training phase. A heat map showing the average reproduction rate for each ML algorithm with $d = 3$, as expressed in Eq. (1), is shown in Fig. 7(a). Analogously, a heatmap showing the average reproduction rate for each ML algorithm with $d = 4$, as expressed in Eq. (1), is shown in Fig. 7(b). A heat map showing the average recall for each ML with $d = 5$, as expressed in Eq. (1), is shown in Fig. 7(c). As the amount of training data increased with the training period s, the amount of training data for each algorithm increased, which led to an improvement in accuracy. In particular, high accuracy was observed for SVM and ANN. The score improved as the lag value f increased, but the recall decreased notably when lag Δt was set to three values. We believe that the reason for the lower recall is that the detected leaves fell off after 3 d or were pruned by farmers, resulting in more FP results.

5.3 Disease forecast per sensor type

We performed a sensor-by-sensor comparison of all the sensors to which we applied the feature engineering methodology proposed in this study. Before the comparison, Figs. 8(a)–8(c) show the average recall for d, Δt, and s in all ML algorithms. The items to be compared among sensors were selected from these figures. Figure 8 shows that the highest recall was achieved by

method	lag Δt days	Training period[S Days] 3	4	5	6
LightGBM	0	0.566	0.593	0.555	0.544
	1	0.586	0.650	0.634	0.346
	2	0.412	0.587	0.468	0.555
	3	0.297	0.667	0.608	0.272
	4	0.520	0.311	0.324	0.491
	5	0.627	0.645	0.569	0.618
XGBoost	0	0.621	0.970	0.830	0.740
	1	0.879	0.859	0.770	1.000
	2	0.816	0.884	0.968	1.000
	3	0.718	0.741	0.705	0.758
	4	0.851	0.640	0.585	0.928
	5	0.462	0.958	0.995	0.986
SVM	0	0.701	0.903	0.696	0.828
	1	0.636	0.851	0.805	0.593
	2	0.742	0.984	0.697	0.990
	3	0.781	0.886	0.883	0.651
	4	0.636	0.646	0.778	0.875
	5	0.701	0.878	0.810	0.860
ANN	0	0.815	0.930	0.887	0.737
	1	0.890	0.933	0.769	0.513
	2	0.917	0.914	0.941	0.899
	3	0.764	0.935	0.641	0.738
	4	0.962	0.979	0.641	0.928
	5	0.727	0.965	0.986	0.976
RF	0	0.577	0.592	0.649	0.648
	1	0.646	0.870	0.678	0.502
	2	0.540	0.763	0.654	0.764
	3	0.213	0.433	0.484	0.196
	4	0.431	0.446	0.416	0.486
	5	0.477	0.842	0.689	0.655

(a)

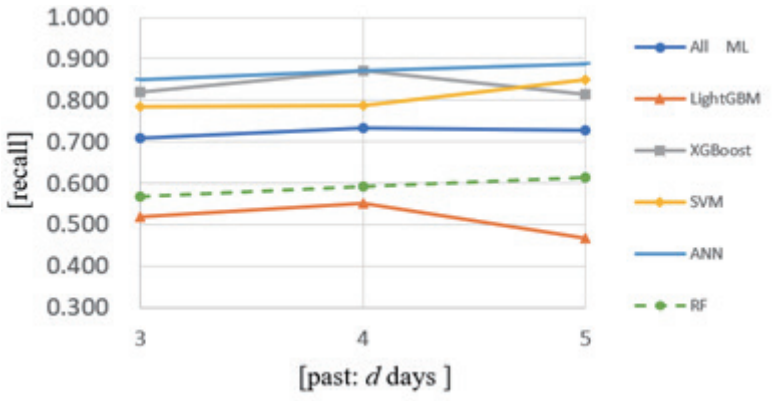
method	lag Δt days	Training period[S Days] 3	4	5	6
LightGBM	0	0.596	0.546	0.408	0.629
	1	0.600	0.484	0.526	0.370
	2	0.493	0.682	0.807	0.556
	3	0.312	0.376	0.470	0.267
	4	0.482	0.565	0.541	0.539
	5	0.692	0.664	0.734	0.888
XGBoost	0	0.937	0.889	0.995	0.973
	1	0.533	0.845	0.805	0.813
	2	0.928	1.000	0.966	0.808
	3	0.642	0.859	1.000	0.967
	4	0.655	0.651	0.823	1.000
	5	0.932	0.987	0.911	1.000
SVM	0	0.971	0.931	0.950	0.781
	1	0.658	0.809	0.763	0.756
	2	0.738	0.971	0.945	0.927
	3	0.609	0.807	0.353	0.497
	4	0.792	0.725	0.883	0.533
	5	0.895	0.818	0.782	0.983
ANN	0	0.901	0.856	0.980	0.948
	1	0.924	0.978	0.723	0.679
	2	0.845	0.961	0.965	0.961
	3	0.914	0.780	0.972	0.442
	4	0.894	0.782	0.932	0.632
	5	0.872	1.000	0.984	1.000
RF	0	0.706	0.731	0.726	0.954
	1	0.603	0.578	0.574	0.538
	2	0.590	0.567	0.680	0.761
	3	0.365	0.352	0.303	0.180
	4	0.513	0.478	0.552	0.513
	5	0.619	0.753	0.693	0.871

(b)

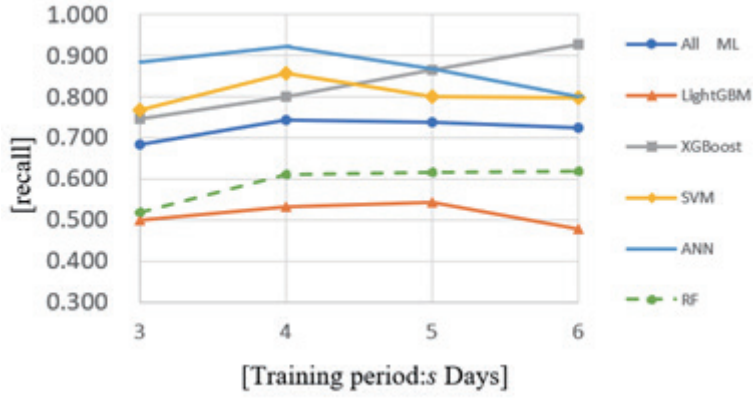
method	lag Δt days	Training period[S Days] 3	4	5	6
LightGBM	0	0.526	0.444	0.690	0.528
	1	0.504	0.280	0.369	0.296
	2	0.393	0.475	0.633	0.453
	3	0.370	0.516	0.467	0.502
	4	0.396	0.459	0.449	0.239
	5	0.604	0.628	0.515	0.499
XGBoost	0	0.920	0.680	0.890	1.000
	1	0.744	0.684	0.848	0.848
	2	0.653	0.775	0.864	1.000
	3	0.847	0.820	0.863	0.951
	4	0.530	0.484	0.970	0.913
	5	0.790	0.702	0.798	1.000
SVM	0	0.797	0.803	0.711	0.771
	1	0.801	0.730	0.671	0.887
	2	0.752	0.956	1.000	0.880
	3	0.883	1.000	0.851	0.825
	4	0.829	0.889	1.000	0.800
	5	0.924	0.838	0.850	0.934
ANN	0	0.973	0.871	0.912	0.969
	1	0.761	0.987	0.543	0.674
	2	0.833	0.961	0.947	0.789
	3	0.947	0.932	0.923	0.917
	4	1.000	0.979	0.953	0.646
	5	1.000	0.864	0.953	0.980
RF	0	0.526	0.551	0.601	0.867
	1	0.667	0.921	0.884	0.966
	2	0.408	0.582	0.669	0.415
	3	0.479	0.325	0.539	0.575
	4	0.413	0.526	0.559	0.478
	5	0.559	0.687	0.746	0.791

(c)

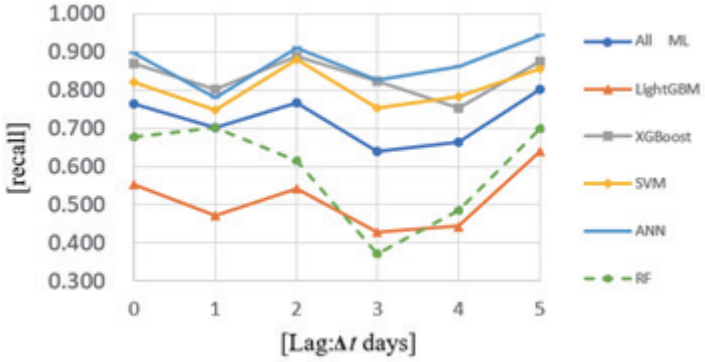
Fig. 7. (Color online) Heat maps of average recall by lag feature: (a) the last 3 d, (b) the last 4 d, and (c) the last 5 d.



(a)



(b)



(c)

Fig. 8. (Color online) Relationship between recall and each parameter: (a) past application period d, (b) lag days Δt, and (c) training period s.

ANN when $d = 5$. Identically, in Fig. 8(b), ANN had the highest score when $\Delta t = 5$. However, because there was variation among the algorithms in Fig. 8(c), the comparison of all sensors was verified when s was varied.

Figure 9 shows the average recall trained with all sensors and a single sensor. The average recall of all sensors was the highest. In the case of carbon dioxide only, the average recall was

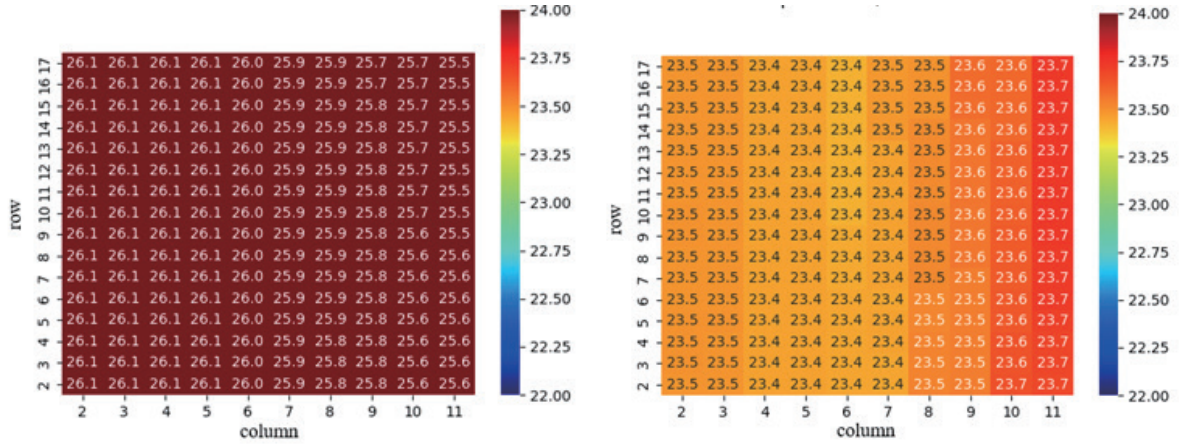
Sensor	Training period[s Days]				Average
	3	4	5	6	
All Sensors	1.000	0.864	0.953	0.980	0.949
Temperture	0.675	0.992	1.000	1.000	0.917
Humidity	0.826	0.793	0.745	0.980	0.836
HD	0.542	0.605	0.567	0.374	0.522
CO_2	0.981	0.922	0.939	0.940	0.946
UV	0.488	0.681	0.659	0.773	0.650
WS	0.717	0.694	0.511	0.736	0.665
Illuminance	0.675	0.667	0.566	0.458	0.592

Fig. 9. (Color online) Recall by environmental sensor.

comparable to that of all sensors. Carbon dioxide was easily understood in terms of photosynthetic activity in plants. On the other hand, while the temperature values alone were also accurate, we focused on the farm temperature heatmap used in the forecast to deepen the discussion. Figure 10(a) shows the observed farm temperature heat map on the forecast day when the wind speed was at its minimum (0.9 m/s). Figure 10(b) shows the observed farm temperature heat map on the forecast day when the wind speed was maximum (4.0 m/s). The variation in the temperature observed on the same day was small, regardless of the wind speed. Therefore, all the observation points were detected as diseased, resulting in a recall of 1.0. We hypothesize that the low recall of LightGBM and RF is due to the underperformance of the leafwise algorithm owing to small weather changes caused by the short training period. LightGBM and RF exhibited low overall recall. However, these ML models allowed several parameters to be set during training. We expected the recall score to increase by changing the parameters at each training cycle.

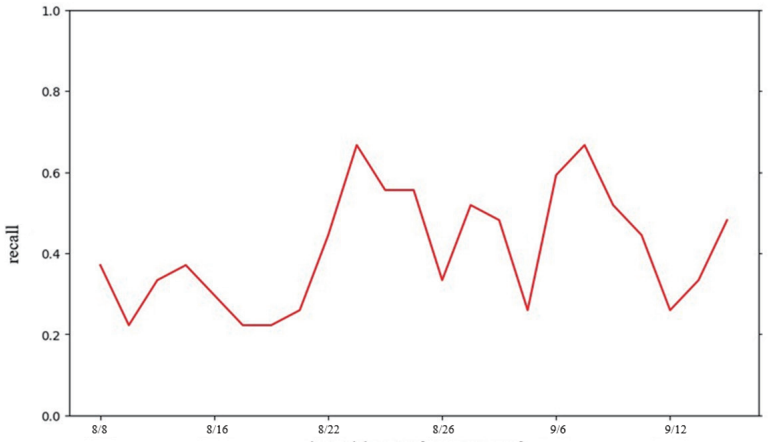
5.4 Duration of training dataset

We validated the case using a larger dataset than the 6 days fixed evaluation dates and increased amounts of training data. Figure 11(a) shows the change in recall when the long-term data were applied to the training data. The test used data from September 14 to forecast 5 d later. The vertical axis represents the recall. The month and day on the horizontal axis represent the starting dates of training. The end date of the training data for both tests was September 13. In other words, the amount of training data increased as one moved to the left along the horizontal axis. The recall decreased from September 12 and August 29 to September 4 and August 26, when rainfall occurred on the same day after the observation time. In the case of the forecast after 5 d, this decrease may be because the sensing agricultural robot's camera could not observe the disease site owing to leaf drop after 5 d, resulting in a shift in the location of the disease. Figure 11(b) shows the change in forecast recall after 1 d, which is assumed to be a day with a low likelihood of defoliation. As shown in Fig. 11(b), the degree of decrease in recall on rainy

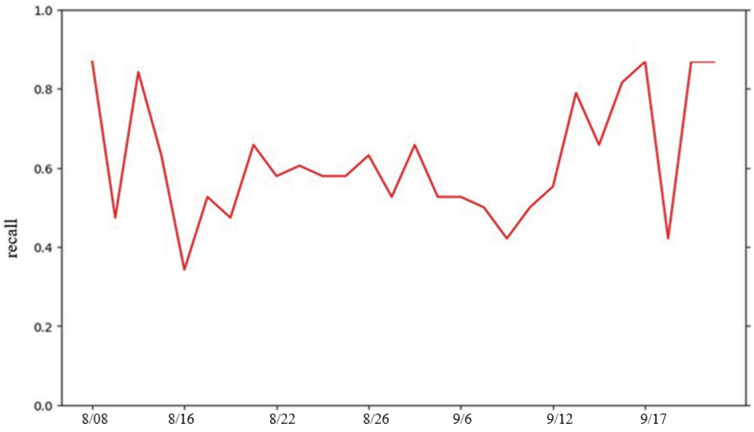


(a) (b)

Fig. 10. (Color online) Temperature heat maps: (a) August 17, wind speed: 0.9 m/s and (b) August 19, wind speed: 4.0 m/s.



(a)



(b)

Fig. 11. (Color online) Variation in recall ratio with long-term datasets: (a) evaluation date: September 14, 2022 and (b) evaluation date: September 22, 2022.

days was small. These results indicate that defoliation and pruning should be considered when forecasting diseases on the basis of the underside of the leaves. In addition, as the amount of training data increased, the recall did not improve, particularly when the training was conducted on data from more than one month after the evaluation date, owing to the accumulation of datasets with low recall.

6. Conclusions

We investigated a method of forecasting diseases five days in advance using supervised ML by photographing the undersides of fruit tree leaves and sensing the environment at the point of photography to forecast diseases using sensing agricultural robots to patrol and observe farms. We developed a disease classifier with a performance of 0.90 accuracy and 0.91 recall using 330 leaf underside images collected from photographs taken by the robot-mounted camera. The results of the disease classifier were used as binary classifications for the objective variable of disease prediction. The final dataset used for learning was created by adding statistical features calculated from the observed values of the previous few days to the raw data. The realization of disease prediction was achieved by lagging the acquisition date of the objective and explanatory variables. SMOTE + ENN was effective for diseased data, which was the observed unbalanced data.

The famer requests a forecast for two days ahead. The best recall after two days was 0.789, which was achieved using an ANN trained on 13 days of dataset ($\Delta t = 2$, $s = 6$, $d = 5$). However, the recall after five days was 0.980, which was achieved using an ANN trained on 16 days of dataset ($\Delta t = 5$, $s = 6$, $d = 5$). This exceeded the farmer's request. As a result of the comparison of all sensor contributions, we found that the results obtained using the CO_2 sensor alone (recall: 0.946) were nearly identical to those obtained using all sensors (recall: 0.949). In the case of disease forecast using data collected over a long term of several months, it was proven that the recall decreases when there is a period of unobserved data due to rainfall, e.g., 7 d, during agricultural work. The supervised ML using a dataset collected over a long period of time with unobserved periods due to, for example, various natural factors, rainy days, typhoons, and pesticide spraying events would not be effective for agricultural support. Thus, we believe that the proposed method using data collected over a short period of time is an effective solution.

However, verification using other farms and agricultural sensors has not yet been conducted. Therefore, an optimal sensor for forecasting is yet to be identified. In the future, it will be necessary to verify the dataset using other farms with different surrounding environments. In addition, the plant used in this study was mulberry, and as an application of the method, it is possible to apply it to other fruit trees and to verify the prediction of diseases using classifiers that focus on specific diseases, such as rust and powdery mildew.

Acknowledgments

The development and support of the small agricultural robot were by Pentalink Co. Data collection in the mulberry fields was made possible by Kogure Farms.

References

1 S. Sieglinde, Sapkota, T. B. Sapkota, C. Jordan, J. Chamberlin, C. M. Cox, S. Gameda, M. L. Jat, P. Marenya, K. A. Mottaleb, C. Negra, K. Senthilkumar, T. S. Sida, U. Singh, Z. P. Stewart, K. Tesfaye, and B. Govaerts: Nat. Sustainability **6** (2023) 1. https://doi.org/10.1038/s41893-023-01166-w
2 A. Nurzaman, D. Debashis, and H. Iftekhar: IEEE Internet Things J. **5** (2018) 4890. https://doi.org/10.1109/JIOT.2018.2879579
3 J. Champ, A. Mora-Fallas, H. Goëau, E. Mata-Montero, P. Bonnet, and A. Joly: Appl. Plant Sci. **87** (2020) e11373. https://doi.org/10.1002/aps3.11373
4 Ministry of Agriculture Forestry and Fisheries of Japan: https://www.maff.go.jp/j/seisan/kankyo/hozen_type/h_sehi_kizyun/aki3.html (accessed June 2024).
5 H. Iyatomi: Brain Neural Networks. **26** (2019) 123. https://doi.org/10.3902/jnns.26.123
6 K. M. Esmail, A. Fahad, A. Sultan, and A. Sultan : Alexandria Eng. J. **60** (2021) 4423. https://doi.org/10.1016/j.aej.2021.03.009
7 H. Wan, Z. Lu, W. Qi, and Y. Chen: Proc. 4th Int. Conf. Machine Learning Soft Computing (2022) 5–9. https://doi.org/10.1145/3380688.3380697
8 K. Kirkpatrick: Commun. ACM **62** (2019) 14. https://doi.org/10.1145/3297805
9 A. Mary: Arabidopsis Protocols **82** (1998) 19. https://doi.org/10.1385/0-89603-391-0:19
10 J. K. M. Brown and M. S. Hovmøller: Am. Assoc. Adv. Sci. **5581** (2002) 537. https://doi.org/10.1126/science.10726
11 W. Shaozu, F. Xiucai, Z. Ying, S. Lei, J. Jianfu, and L. Chonghuai: Acta Bot. Boreal.-Occident Sin. **43** (2023) 1352. https://doi.org/10.7606/j.issn.1000-4025.2023.8.1352
12 WAGRI: https://wagri.net/aboutwagri (accessed July 2024).
13 T. Wakabayashi: JATAFF J. **7** (2019) 34. https://agriknowledge.affrc.go.jp/RN/2010928564 (accessed June 2024).
14 Plant protection: https://www.jppn.ne.jp/jpp/s_mokuji/20190610.pdf (accessed July 2023).
15 G. Fenu and F. M. Malloci: Proc. 3rd Int. Conf. Big Data Research (2019) 76–82. https://doi.org/10.1145/3372454.3372474
16 F. Gianni and M. F. Maridina: Big Data Cognit. Comput. **5(1)** (2021) 2. https://doi.org/10.3390/bdcc5010002
17 N. Petrellis: Proc. 21st PanHellenic Conf. Informatics **1** (2017) 1–6. https://doi.org/10.1145/3139367.313936
18 S. A. A. Qadri, N. Huang, T. M. Wani, and S. A. Bhat: IEEE 13th Int. Conf. Control System, Computing and Engineering (ICCSCE). (2023) 155–160. https://doi.org/10.1109/ICCSCE58721.2023.10237169
19 L. Lou, J. Liu, Z. Yang, X. Zhou, and Z. Yin: Proc. 2022 6th Int. Conf. Computer Science and Artificial Intelligence (2022) 7–12. https://doi.org/10.1145/3577530.3577532
20 Z. Shi, C. Wang, and L. Zhao: Proc. the 6th Int. Conf. Computer Science and Application Engineering (2022) 1–6. https://doi.org/10.1145/3565387.3565399
21 IPSJ SIG Technical Report: https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_action_common_download&item_id=182477&item_no=1&attribute_id=1&file_no=1 (accessed July 2024).
22 Y. Chenghai: Engineering **6** (2022) 528. https://doi.org/10.1016/j.eng.2019.10.015
23 A. Sugiura: Jpn. J. Pestic. Sci. **45** (2020) 146. https://doi.org/10.1584/jpestics.W20-22
24 H. Fernandes, E. C. M. Polania, A. P. Garcia, O. B. Mendonza, and D. Albiero: Rev. Ciênc. Agron. **51** (2020) 1. https://doi.org/10.5935/1806-6690.20200092
25 L. Grimstad: IFAC-PapersOnLine **50** (2017) 4588. https://doi.org/10.1016/j.ifacol.2017.08.1005
26 P. Gonzalez-De-Santos, R. Fernández, D. Sepúlveda, E. Navas, and M. A. Chang: Food Secur. **6** (2020) 73. http://dx.doi.org/10.5772/intechopen.90683
27 S. J. LeVoir, P. A. Farley, T. Sun, and C. Xu: IEEE Open J. Ind. Appl. **1** (2020) 74. https://doi.org/10.1109/OJIA.2020.3015253
28 G. Quaglia, C. Visconte, L. Carbonari, A. Botta, and P. Cavallone: IEEE Open J. Ind. Appl. **1** (2020) 81. https://doi.org/10.1007/978-3-030-55757-7_6
29 M. R. Tammina, K. Sumana, P. P. Singh, T. R. V. Lakshmi, and S. D. Pande: Microbial Data Intelligence and Computational Techniques for Sustainable Computing. (2024) 25. https://doi.org/10.1007/978-981-99-9621-6_2
30 K. Terada, M. Endo, T. Kikuchi, and S. Ohno: Int. J. Inf. Soc. (2023) 135. http://www.infsoc.org/journal/vol14/14-3
31 M. Honda, K. Tatsumi, and M. Nakagawa: Jpn. Soc. Agric. Inf. **30** (2021) 96. https://doi.org/10.3173/air.30.96
32 G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard: ACM SIGKDD Explorations Newslett. **6** (2004) 20. https://doi.org/10.1145/1007730.1007735

33 T. Chen and C. Guestrin: Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (2016) 785–794. https://doi.org/10.1145/2939672.2939785
34 G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu: Proc. the 31st Int. Conf. Neural Information Processing Systems (2017) 3149–3157. https://dl.acm.org/doi/pdf/10.5555/3294996.3295074
35 C. Wu, X. Xue, and Y. Song: Proc. the 4th Int. Conf. Big Data Engineering (2022) 122–126. https://doi.org/10.1145/3538950.35389
36 T. Ying and W. Jun: IEEE Trans. Knowl. Data Eng. **16** (2004) 385. https://doi.org/10.1109/TKDE.2004.1269664
37 L. Breiman: Random For. **45** (2001) 5. https://doi.org/10.1023/A:1010933404324