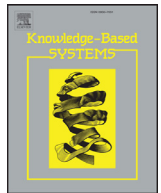


Multi-modal sliding window-based support vector regression for predicting plant water stress

メタデータ	言語: en 出版者: Elsevier B.V. 公開日: 2018-01-11 キーワード (Ja): キーワード (En): 作成者: Kaneda, Yukimasa, Shibata, Shun, Mineno, Hiroshi メールアドレス: 所属:
URL	http://hdl.handle.net/10297/00024378



Multi-modal sliding window-based support vector regression for predicting plant water stress



Yukimasa Kaneda^a, Shun Shibata^a, Hiroshi Mineno^{b,c,*}

^a Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka 432-8011, Japan

^b College of Informatics, Academic Institute, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka 432-8011, Japan

^c JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan

ARTICLE INFO

Article history:

Received 16 March 2017

Revised 20 July 2017

Accepted 23 July 2017

Available online 24 July 2017

Keywords:

Water stress

Image processing

Deep neural network

Support vector regression

Ensemble learning

ABSTRACT

Information communication technology (ICT) is required in the field of agriculture to solve problems arising because of the aging of farmers and shortage of heirs. In particular, environmental sensors and cameras are widely used in existing agricultural support systems for easy data collection. Although the traditional purpose of these systems is naive monitoring and controlling of the environment, the propagation of advanced cultivation is now expected by applying the data to machine learning and data mining technologies. Therefore, we propose a novel multi-modal sliding window-based support vector regression (multi-modal SW-SVR) method for accurate prediction of complicated water stress, which is a plant status, from two data types, namely environmental and plant image data. The proposed method includes two methodologies, SW-SVR and deep neural network (DNN) as a multi-modal feature extractor for SW-SVR. SW-SVR, which we proposed previously, is a suitable learning method for data with time-dependent characteristics, such as plant status. Moreover, we propose a new image feature, remarkable moving objects detected by adjacent optical flow (ROAF), to enable DNN to extract essential features easily for predicting water stress. Compared with existing regression models and features, the proposed multi-modal SW-SVR with ROAF demonstrates more precise and stable water stress prediction.

© 2017 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Sensing technology is becoming more widespread and sophisticated, and many studies on artificial intelligence have demonstrated the propagation of sophisticated intelligence by detailed analysis of the data. In the field of agriculture, many studies on the propagation of advanced cultivation have been conducted to solve several problems arising because of the aging of farmers and shortage of heirs. These studies quantify and analyze complicated plant statuses from various data, such as environmental data, plant image data, and growth status, and farmer's decisions based on experience and intuition can be reproduced [1–4]. In particular, the propagation of stress cultivation based on water stress is strongly expected to improve profit for fruits with high sugar contents.

The mechanism for producing sweet fruits by stress cultivation is shown in Fig. 1. Water stress on plants causes them to expel water from fruit and increase its sugar content ratio [5]. Therefore, fruit size decreases, but sugar content increases. Although the

amount of water and sugar content in fruit is inversely proportional, excessive water limitation leads to the withering of plants. Therefore, by determining the ease-of-withering from water stress, and irrigating plants shortly before they die, the sugar content can be increased by repeating the procedure shown in Fig. 1. Meanwhile, if plants are subject to considerable withering at least once, they will die despite being irrigated. It is important to predict future water stress and irrigate plants prior to considerable withering. Accordingly, the accurate prediction of water stress enables cultivation of fruits with high sugar contents.

Important data in rapid prediction of water stress includes environmental data, such as temperature, humidity, solar quantity, and plant fertility, because it affects the plants' physiological processes, which causes water stress [6]. Water stress occurs particularly when the amount of transpiration from the leaves exceeds that of the water supplied from the root. Given that the environmental data is strongly related to transpiration, it is also indirectly related to water stress. Other data in which water stress appears remarkably is plant image data. Fig. 2 shows example images of plants under water stress. The leaves after exposure to water stress shrivel up and droop compared to those before water stress. Thus, water stress is expressed directly in the plant image data because

* Corresponding author.

E-mail addresses: kaneda@minelab.jp (Y. Kaneda), mineno@inf.shizuoka.ac.jp (H. Mineno).

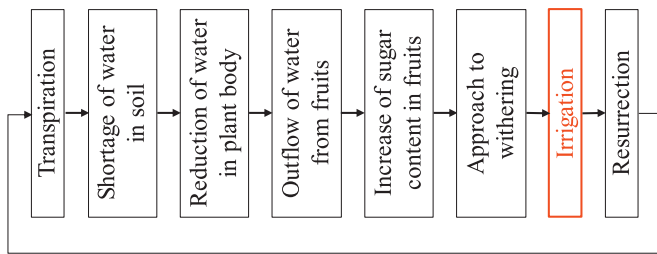


Fig. 1. Mechanism of increasing sugar content.

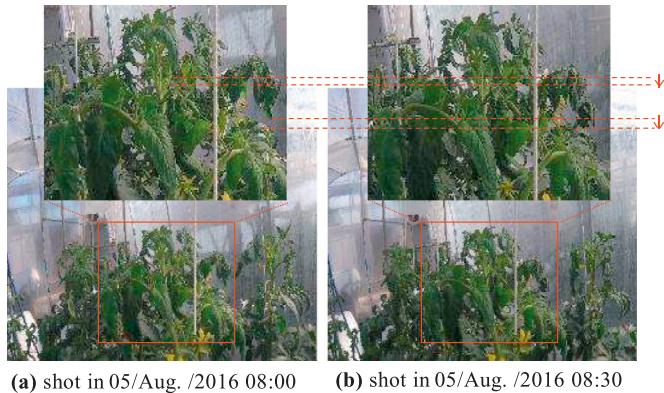


Fig. 2. Example of variation in plant under water stress.

leaves wilt significantly owing to water stress. Given that the two features complement each other's insufficiencies, complicated water stress is expressed multilaterally.

Meanwhile, there are two issues in predicting water stress accurately from the two kinds of data. First, given that water stress variation depends strongly on the complicated change in the natural environment over time, water stress is imbalanced data that differs considerably in the amount of data for each characteristic. In the field of agriculture, previous studies have dealt with imbalanced data such as water stress, plant disease, and environmental factors [7–9]. Meanwhile, the majority of methods are restricted to classification problems, and limited studies deal with the regression problem as water stress [9]. Furthermore, previous methods have addressed the imbalance only during the training phase for over-sampling, under-sampling, and weighting training data. However, according to our previous research [10], training data that constructs models should change according to the characteristic variation with time of the testing phase. For example, in ensemble learning, the weights to aggregate each model built from different kinds of training data should depend on the variation in the test data. The second problem is image feature extraction from complicated plant image data. Recently, convolutional neural network (CNN), which is a type of deep neural network (DNN), has been widely used for several applications, and surpassed human intelligence in some computer vision tasks, such as face recognition, object detection, and object recognition. Furthermore, CNN has demonstrated state-of-the-art results with respect to image based plant phenotyping [11]. Therefore, CNN can be expected to extract image features related to water stress appropriately. CNN extracts essential features in complicated image data because it learns how to extract features and main problems with non-linear processing for image recognition. However, features required for water stress prediction are not always extracted correctly despite using CNN as a feature extractor. This failure is attributed to unnecessary information occupying a large proportion compared with necessary information for water stress prediction in plant image data acquired in the greenhouse. Consequently, the unnecessary in-

formation must be removed before inputting CNN for easy extraction of features related to water stress.

In this paper, we propose a novel learning methodology for predicting plant water stress called multi-modal sliding window-based support vector regression (multi-modal SW-SVR). SW-SVR, which we proposed previously [12], specializes in predicting more accurately imbalanced data with time-dependent characteristics, such as plant status, than other ensemble regression methods, such as gradient boosting regression, AdaBoost, bagging decision tree, and random forest. The basic theories comprise a specific data prediction based on data extraction and newfangled weighting that determines weights in ensemble learning according to time-dependent characteristics. The weighting of each specialized model supplements the mutual strength and weaknesses with each other. Furthermore, multi-modal SW-SVR is an SW-SVR where DNN with a novel image feature is applied to extract essential multi-modal features from the two data types. Moreover, we propose an image feature to enables DNN to extract essential features easily to predict water stress; this feature is called remarkable moving objects detected by adjacent optical flow (ROAF), which reproduces the same perspective as farmers on original image data by extracting the plant status of wilted leaves with water stress based on plant wilting motion.

The remainder of this paper is organized as follows. Section 2 presents related work and our previously proposed method in predicting water stress. The proposed methods are described in detail in Section 3. Section 4 presents the experiments for evaluation of the proposed methods. Finally, the paper concludes in Section 5.

2. Preliminaries

2.1. Propagation of stress cultivation

Previously, conventional agricultural support systems collected environmental data using sensor network technology to monitor environmental data and control the environment naively [13]. Recently, further utilization of information communication technology (ICT) for agricultural support has been livelier to propagate advanced cultivation of farmers. In particular, the realization of the watering system dependent on water stress is expected to cultivate fruits with high sugar content. Many studies regarding water stress prediction have been conducted. Conventionally, water stress is measured on the basis of water potential, which indicates the moisture retention ability of a plant. Therefore, water potential is used as an indicator of water stress, and measuring water potential yields an accurate value of water stress [14]. However, water potential is not obtained in real-time because the measurement methods are based on destructive testing. To measure water stress in real-time, water stress prediction using sensor data, such as environmental data, plant image data, and growth status, are noted. Approaches based on growth status, such as stem-diameter and natural frequency of leaves, predict water stress accurately because these data appear to fluctuate according to water stress [1,2]. Furthermore, the stem-diameter approach enables water stress to be measured with non-destructive and non-contact methods continuously using laser displacement sensors that measure the approximate shape of objects based on laser light injected by the sensor. However, it is difficult to apply these approaches practically because these measuring devices are expensive and require professional knowledge for correct measurement.

On the other hand, the approaches based on environmental data and plant image data are measured easily [3,4] because the sensors and cameras required for measurement are generally inexpensive, and expert knowledge is unnecessary for installation and measurement. Furthermore, the approaches measure the data

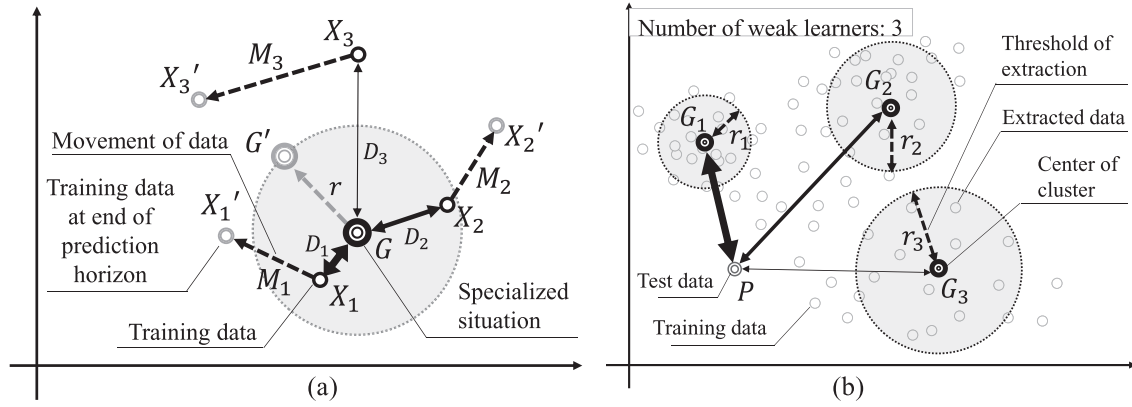


Fig. 3. Processing outline of SW-SVR, (a) D-SDC that extracts training data based on movement of the specialized situation, (b) dynamic weighting depending on variation in test data.

in a non-destructive and non-contact manner. However, compared with the approaches based on growth status, it is difficult to predict complicated water stress that changes based on various factors from only external factors, such as environmental data and plant image data. In other words, there is a trade-off between prediction accuracy and ease-of-measurement.

We aim for a balance in this relationship and predict water stress accurately from environmental data and plant image data using multi-modal SW-SVR based on ingenious machine learning technologies and an image feature. Multi-modal SW-SVR supplements the problem related to prediction accuracy caused by using only an external factor that is easy to collect and improves practicality for predicting water stress compared with conventional approaches. Furthermore, to our knowledge, this is the first-time water stress is predicted by combining environmental data and plant image data using DNN.

2.2. Sliding window-based support vector regression

Previously, we proposed a new methodology called SW-SVR for predicting imbalanced data with time-dependent characteristics, as shown in Algorithm 1. The basic theories comprise specific data prediction based on data extraction and a newfangled weighting that determines weights in ensemble learning depending on test data.

First, SW-SVR builds plural linear support vector regression (SVR), specialized for each representative situation, such as different seasons and climates. The specialized situations are defined by centers of clusters classified by k-means [15], instead of random sampling, so that the specialized situations can represent more various situations. The specialized SVRs are built based on dynamic-short distance data collection (D-SDC) that extracts effective data for specific data prediction by taking into account movement, which is the feature variation during prediction horizons. Extracted training data S by D-SDC is based on r , which is a movement of a specialized situation, as shown as follows:

$$S_t = \{(\mathbf{x}_i, y_i) \mid \|\mathbf{G}_t - \mathbf{x}_i\| < r_t\}, \quad r_t = \|\mathbf{G}_t - \mathbf{G}'_t\|,$$

where $\mathbf{x}$ is the feature set of the training data, y is the dependent variable of the training data, $\mathbf{G}$ is a specialized situation, and $\mathbf{G}'$ is a specialized situation after prediction horizons. D-SDC extracts training data whose norm from a specialized situation is shorter than movement r , as shown in Fig. 3(a). D-SDC is based on movement r because we consider that the amount of data required for predicting specific data depends on the characteristics of the specific data. In particular, more training data is necessary to specialize for a situation that changes dramatically with time. Meanwhile,

movement r is unknown because $\mathbf{G}'$ is not observed at the time. Therefore, D-SDC estimates movement r based on movements in training data, similar to a specialized situation by weighted average, where the weights are reciprocals of the norms between a specialized situation and each training data, as follows:

$$r_t = \|\mathbf{G}_t - \mathbf{G}'_t\| \approx \frac{\sum_{i=1}^N w_i \|\mathbf{x}_i - \mathbf{x}'_{ii}\|}{\sum_{i=1}^N w_i} \quad \text{where } w_i = \frac{1}{\|\mathbf{G}_t - \mathbf{x}_i\|^p},$$

where N is the number of training data, and p is a weighted parameter. Then, SW-SVR builds plural linear SVRs as specialized models based on the extracted data. Each specialized model accurately predicts the data similar to the specialized situation.

Then, conclusively predicted values of SW-SVR take characteristic variation with time of test data into account. In general ensemble learning for regression, the weights to integrate each model are determined completely in the training stage. However, SW-SVR dynamically determines the weights for each prediction repeatedly, and the weights are decided by the similarity between the test data and each specialized situation in each specialized model as shown in Fig. 3(b). A final hypothesis of SW-SVR is shown as follows:

$$H(\mathbf{P}) = \frac{\sum_{t=1}^n w_t H_t(\mathbf{P})}{\sum_{t=1}^n w_t} \quad \text{where } w_t = \frac{1}{\|\mathbf{G}_t - \mathbf{P}\|^q},$$

where $\mathbf{P}$ is the test data, n is the number of specialized models, $H(\mathbf{x})$ is a hypothesis of each model, and q is a weighted parameter. Despite characteristic variation with time in the test data, SW-SVR always gives priority to specialized models that are more suitable for predicting test data.

Meanwhile, SW-SVR uses two kinds of feature extraction to map into new feature space to take the presence of noise and non-linear relationships into account: kernel approximation [16] and partial least squares (PLS) regression [17]. However, the procedure assumes input of dense sensor data and is not suitable for image data that involves many features and multidimensionality. Moreover, the feature extraction does not take multi-modal features combining image data and environmental data into account. Therefore, SW-SVR is used to predict future water stress accurately based on environmental data and image data; an alternative to feature extraction for multi-modal features must be considered.

2.3. Deep neural network

DNN is becoming more widespread because it learns how to extract features as well as main problems with non-linear processing, and the deep construction can represent all relationships approximately. DNN is used rapidly because recent research has solved

the gradient vanishing problem, which is a long-standing problem occurring in back propagation: pre-training, dropout, batch normalization, and weight initialization [18–21]. However, when input data is multidimensional data like image, the number of parameters becomes enormous and causes overfitting because all neurons between adjacent layers in DNN are connected. To overcome the overfitting problem, CNN, which is a DNN with smaller learning parameters, is widely used and is highly successful in several computer vision tasks [22,23]. CNN, whose structure is based on a visual cortex of the creatures' brain, binds a neuron in each layer to only neurons of local regions called receptive fields, and not to all neurons of the previous layer; additional parameters in the receptive field are shared to all receptive fields. Therefore, it is easier to prevent overfitting in CNN than in DNN using image data. The receptive field with sharing the parameter is regarded as a process of convoluting a two-dimensional filter with an image. The layer with convolutional operation is called a convolutional layer in CNN. Furthermore, CNN also has pooling layers to reduce computational complexity. Repeating convolutional and pooling layers enables CNN to extract essential features from an image by passing only distinguishable information of the features obtained by the convolution process to the next layer.

However, features required for water stress prediction are not always extracted correctly, despite using CNN as a features extractor, because in plant image data, unnecessary information occupies a large proportion compared with the necessary information for water stress prediction. The necessary information is the only plant status of the wilted leaves with water stress, such as texture, luster, color, and shape. Meanwhile, plant image data also involves considerable unnecessary information, such as background and sunlight. Therefore, we propose a new image feature to enable CNN to easily extract the plant status related to water stress from complicated plant image data. The proposed feature reproduces the same perspective as farmers on original image data by extracting only the plant status of wilted leaves with water stress based on plant wilting motion.

3. Proposed approach

3.1. Plant water stress prediction using multi-modal SW-SVR

We propose a novel plant water stress prediction method using multi-modal SW-SVR to reproduce farmer's cultivation precisely. Our method uses two kinds of data: image data in which water stress variation is expressed directly as plant wilting, and environmental data related to plant transpiration, which is the cause of water stress. Given that the two features complement each other's insufficiencies, complicated water stress is expressed multilaterally.

Multi-modal SW-SVR is an SW-SVR to which DNN with a novel image feature is applied to extract essential multi-modal features from the two data types (Fig. 4). SW-SVR is suitable for data with time-dependent characteristics, such as plant status. SW-SVR, which we proposed previously, builds many models specialized for various situations and aggregates the models effectively according to time-dependent characteristics. Meanwhile, DNN can extract multi-modal features from different data types because non-linear processing is applied to each feature comprehensively [24]. CNN extracts particularly essential features suitable for main problems from image data, unlike the complicated conventional methods that require human expertise. In the proposed multi-modal SW-SVR, multi-modal features are extracted by DNN based on environmental data, and a proposed image feature from which only features related to water stress in plant image data are easily extracted. Specifically, features are first extracted from plant image data in CNN, and then integrated with environmental data in the same network. Then, SW-SVR predicts water stress based on the

multi-modal features while following water stress variations with time.

In the network, an image processing method for motion detection, called optical flow, is applied to image data to extract important features related to water stress in similar plant image data. Optical flow represents motion based on spatiotemporal variation from two image data, and leaves with water stress are clarified by optical flow because water stress in plant image data is expressed as wilting motion. Furthermore, we propose an image feature using plural optical flow, called remarkable moving objects detected by adjacent optical flow (ROAF), which extracts the plant status of the wilted leaves with water stress based on plant wilting motion; this is unlike conventional region extraction methods. Specifically, ROAF reproduces the same perspective as farmers on the original image data.

The system architecture for our method is shown in Fig. 5. We assume that the proposed system is processed on our previously implemented agriculture support system for greenhouses [25]. In the previously proposed system, a network in a greenhouse collects image data and environmental data, and sends control signals to control the equipment. Moreover, in the cloud system, future water stress is predicted, and control signals based on the prediction is sent to the greenhouse. We added cameras to take plant image data, and stem-diameter sensors, to the system. Stem-diameter sensors are only used for labeling of grand-truth for predicting water stress, which is described in detail in the succeeding section. In this study, mainly the water stress prediction methodology in Fig. 5 is mentioned. Finally, we assume that actuators for irrigation, such as pumps, are controlled based on the prediction; then, plants are watered before they die.

3.2. Remarkable moving objects detected by adjacent optical flow

ROAF consists of the following procedures: optical flow (OF) and pooled optical flow (POF). The process of generation for one ROAF and an example of ROAF are shown in Fig. 6(a) and (b), respectively. First, an optical flow, which is an image processing technology, is used to recognize the wilting motion of plants. Optical flow represents motion based on spatiotemporal variation from two image data. In our methods, deep flow [26], which is an optical flow algorithm, is used. Deep flow extracts dense optical flow, and the motion of non-rigid objects can be grasped. Therefore, the plant's motion is detected easily and robustly in plant image data that involves considerable information not related to water stress.

Optical flow expresses the motion including moving leaves due to water stress, from the image data. However, optical flow can extract only the difference between two points, and any movement that occurred previously cannot be considered completely. Meanwhile, even if there is no movement on the leaves at the current time, the water stress on the leaves is given if the leaves withered in the past. Therefore, features related to water stress can also be extracted from leaves that wilted during past fixed periods. Then, we propose the POF as a new optical flow methodology considering a wilting motion that has already occurred. The main methodology of POF is pooling adjacent optical flow. Basic pooling processing in CNN is a non-linear down-sampling methods for one channel, but maximum values for each pixel from the plural optical flow are calculated in our pooling processing. As a result, the new optical flow is generated considering all withering motion that occurred in previous fixed periods. Specifically, POF represents the location of only the leaves subject to water stress. POF is also effective in terms of noise removal because only motions with change over a time are emphasized. The motion counteracts temporal noise and motion on leaves. Given that noise caused mainly by the wind occurs frequently and irregularly on the leaves, it is useful to be able to remove these noises automatically based on

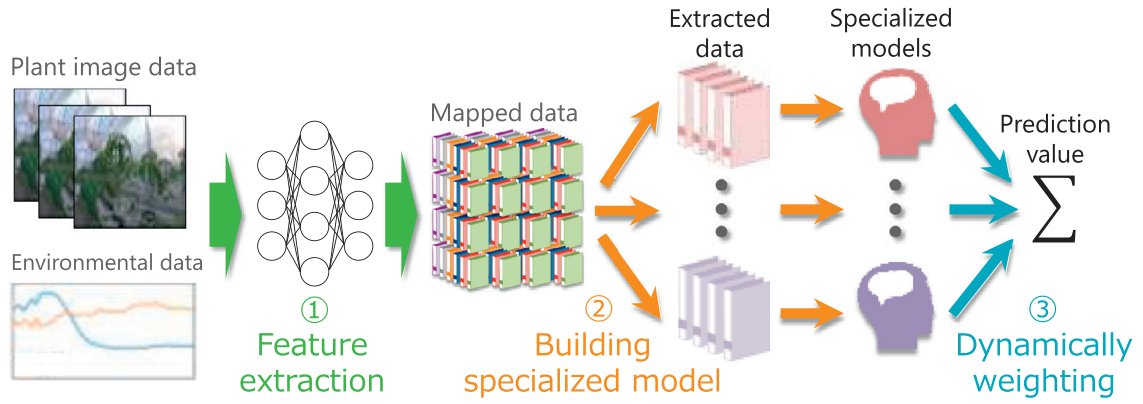


Fig. 4. Overview of multi-modal SW-SVR.

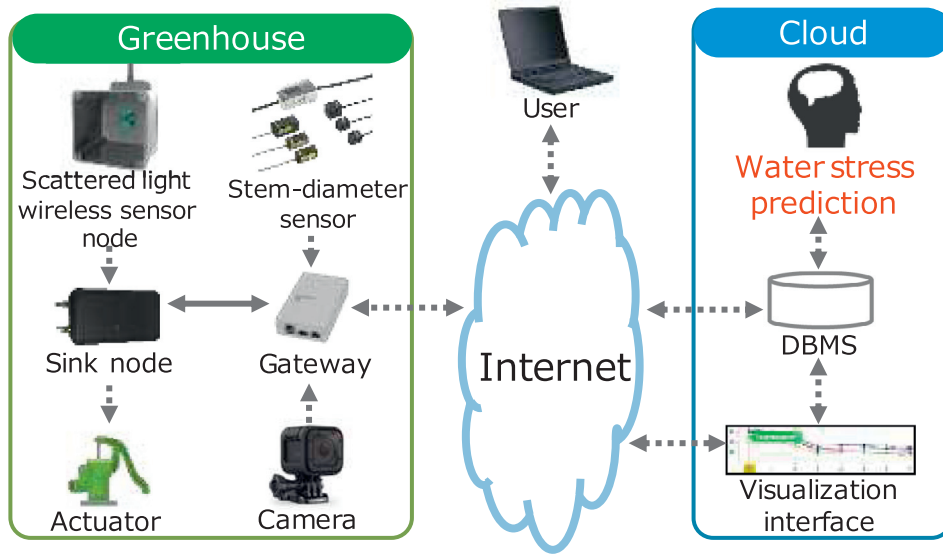


Fig. 5. System architecture.

the relative difference in motion. If the maximum values obtained from POF are very small, on a supposition that water stress cannot occur on the pixel, the pixel is excluded from the extraction target pixels.

To irrigate plants correctly during stress cultivation, farmers have focused on some important plant statuses of leaves, such as texture, luster, color, and shape, which are included in the original image data. Meanwhile, we propose a POF that describes the location of only leaves subject to water stress by combining plural optical flow. Finally, we apply POF to the original image data that includes considerable information to extract the above plant status from only leaves with water stress focused on by POF; the extracted original image data by POF is named ROAF. ROAF is generated by mask processing when the POF is used as a mask image, and only pixels with detected optical flow in POF are extracted in the original image data. ROAF reproduces the same perspective as farmers in the original image data, and CNN can be learned using image features obtained from the viewpoint because when all values of the targeted plural pixels by the kernel in convolutional layer become zero, the information is not transmitted from the pixels to the neurons in the next layer. Specifically, the neurons are not firing. By setting the pixels in which POF is not detected to zero, the pixel information is not transmitted and is not used to update weights in CNN. Consequently, weights in CNN are updated using only the important plant status considered by farmers in the original image data.

3.3. Network architecture for DNN

Fig. 7 and Table 1 show a network architecture for DNN that achieves compatibility between reducing the number of parameters and improving the generalization performance. Considerable image data with labels is generally collected through an internet service, such as ImageNet [27], for problems of general object recognition. Therefore, even if the number of parameters increases owing to the complication of network architecture, DNN can learn correctly from the considerable image data. However, given that there is no plant image data associated with stem-diameter available on the internet, the image data must be generated, and it is difficult to gather many of the plant image data. Meanwhile, fewer data compared to the amount of the parameters leads to overfitting in the learning process. Accordingly, the network focuses on the trade-off relationship between reducing the number of parameters and improving generalization performance.

Contrivances in our DNN are described below. First, initial values for weight parameters are determined based on He initialization [21] for efficient learning in DNN. Initial values are highly important because the progress of learning depends greatly on the initial values. Previously, initial values were defined randomly in general, but He initialization is based on Gaussian distribution and disperses all activation maps. The network is trained efficiently based on moderately diverse data. Next, network A is a general four-layer CNN with batch normalization layers [20] to extract

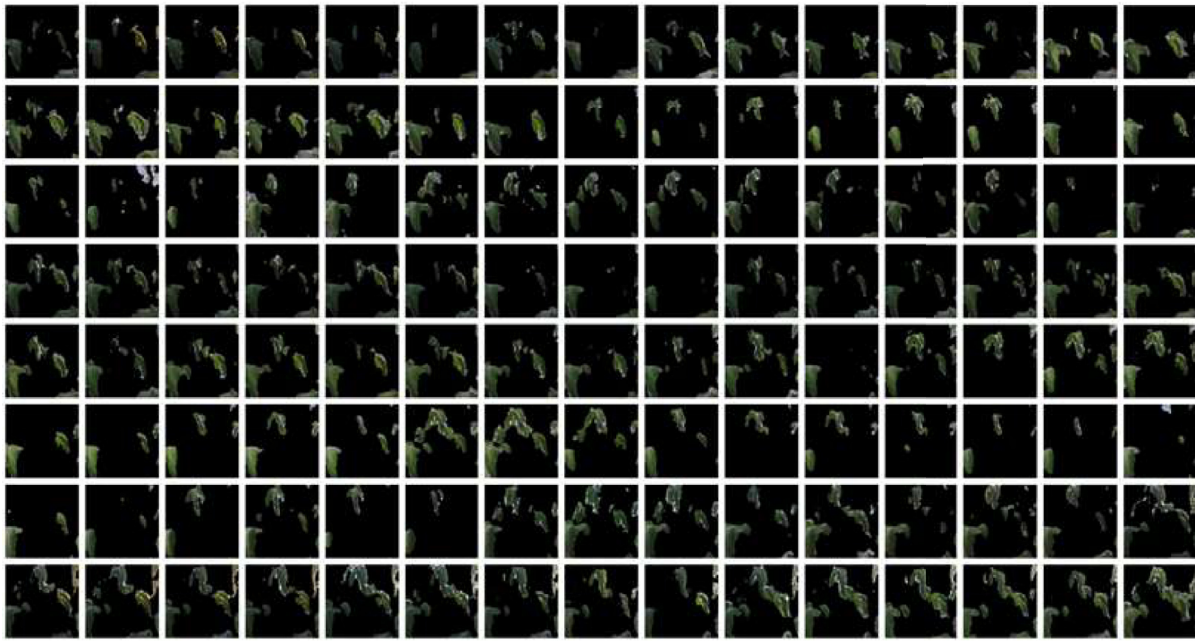
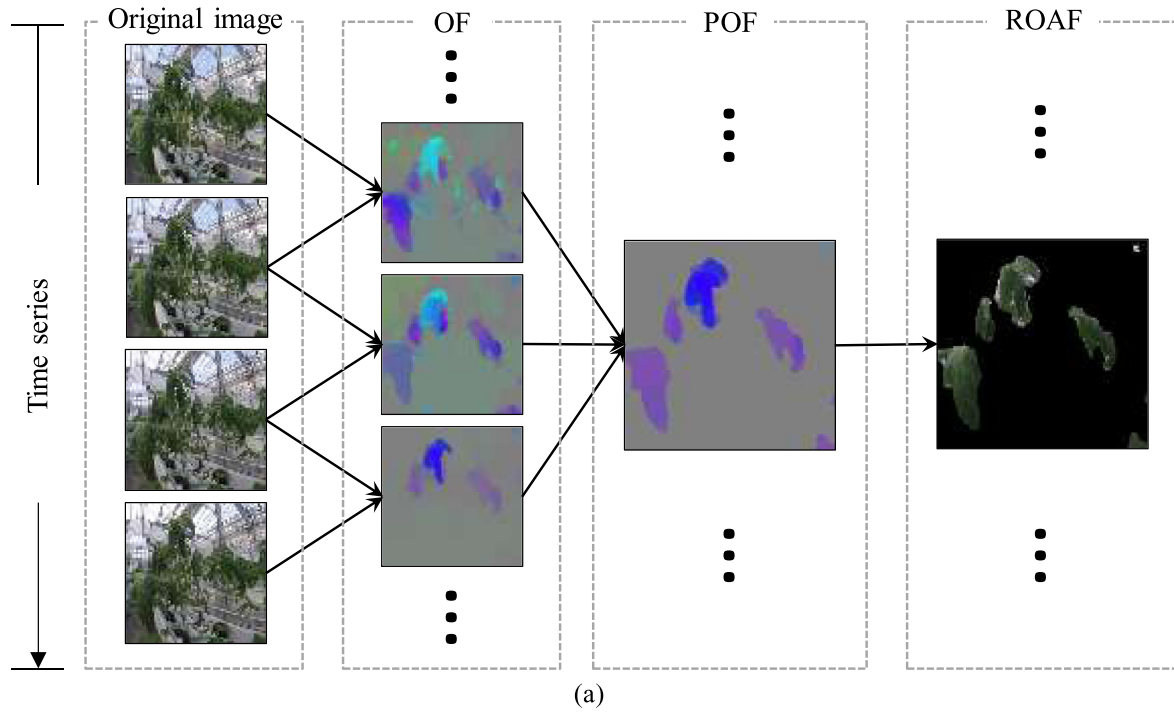


Fig. 6. Feature extraction based on the proposed ROAF, (a) process of generation for one ROAF, (b) example of ROAF.

essential features from ROAF efficiently. Batch normalization is used for acceleratory learning in DNN, and the role of batch normalization is similar to He initialization. Batch normalization adjusts the activation distribution to vary it properly. The kernel size of the convolution layer was limited to three to reduce learning parameters. Then, network B enlarges the number of dimensions of the environmental data by using non-linear conversion in the fully connected layer. Given that environmental data has never had a complicated relationship compared with the image data, one fully connected layer is used to extract essential features from the environmental data. Finally, network C integrates the features based on

the image data and environmental data extracted by networks A and B, respectively, in a 1×1 convolution layer as the fusion layer. By setting the number of sheets of kernels in the 1×1 convolutional layer to less than the number of channels of the inputted data, the new data with fewer channels is outputted. The layer has few learning parameters, but the features between different channels can be integrated efficiently compared with other fusion methods [28]. Finally, the integrated features are inputted to two fully connected layers.

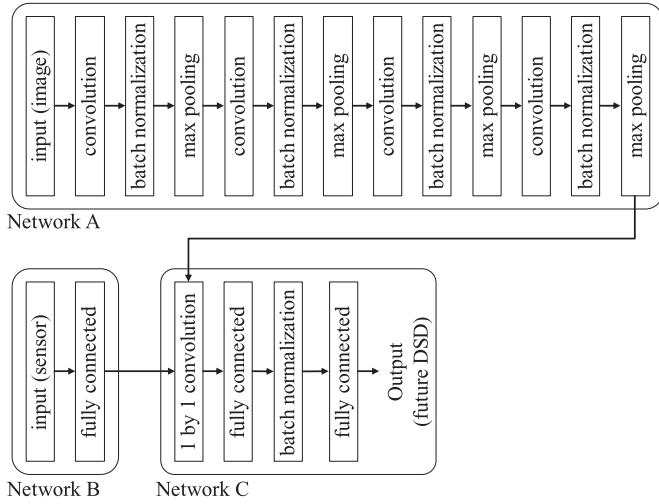


Fig. 7. Network architecture for multi-modal feature extraction.

Table 1

Detail of parameters for network as shown in Fig. 7.

(a) Detail of parameters for network A.				
Layer	Patch	Stride	Map size	Function
data	–	–	$144 \times 144 \times 3$	–
conv1	3×3	2	$72 \times 72 \times 24$	PReLU
norm1	–	–	$72 \times 72 \times 24$	–
pool1	2×2	2	$36 \times 36 \times 24$	–
conv2	3×3	1	$36 \times 36 \times 48$	PReLU
norm2	–	–	$36 \times 36 \times 48$	–
pool2	2×2	2	$18 \times 18 \times 48$	–
conv3	3×3	2	$9 \times 9 \times 96$	PReLU
norm3	–	–	$9 \times 9 \times 96$	–
pool3	2×2	2	$5 \times 5 \times 96$	–
conv4	3×3	2	$3 \times 3 \times 96$	PReLU
norm4	–	–	$3 \times 3 \times 96$	–
pool4	2×2	2	$2 \times 2 \times 96$	–
(b) Detail of parameters for network B.				
Layer	Patch	Stride	Map size	Function
data	–	–	6	–
fc	–	–	384	PReLU
(c) Detail of parameters for network C.				
Layer	Patch	Stride	Map size	Function
data	–	–	$1 \times 384 \times 2$	–
fusion	–	–	384	–
fc1	–	–	64	PReLU
norm	–	–	64	–
fc2	–	–	1	PReLU

3.4. Environmental data related to water stress

Water stress occurs particularly when the amount of transpiration from the leaves exceeds that of the water supplied from the root. Transpiration is related to various environmental data, such as temperature, humidity, solar quantity, and plant fertility. Meanwhile, a collection of data must be easy to use as features because the same environment used measure the features must be constructed in applications. Then, we use scattered light wireless sensor nodes [25] to measure all of these data. These are a cube covered with light shielding frames, with the exception of one surface, and the sensor node can collect temperature, humidity, and two solar quantity types. Silicon photodiodes for sensing solar quantity are installed on and in the sensor node. The silicon photodiode on the sensor node collects direct light, and another silicon photodiode in the sensor node collects scattered light. Given that scattered light is not affected by shadows, it always determines the

overall solar brightness. Therefore, solar quantity in plant communities can be measured as scattered light. Moreover, based on the characteristics of scattered light, the sensor nodes installed on and in the plant community can measure plant fertility because plants grow in the direction of the light. Scattered light in a plant community decreases according to plant growth. Consequently, the ratio of scattered light in plant community to that on plant community indicates continuous non-contact plant fertility.

3.5. Definition of water stress

In our method, stem-diameter is measured as the water stress of the dependent variable during the training phase. Water stress occurs by decreasing the amount of water in the plant; the increase and decrease are revealed explicitly by the thickness of stem-diameter. Then, water stress can be measured based on the stem-diameter. According to the study that evaluated the relationship between stem-diameter and water stress, plants subject to water stress tend to increase the maximum variation in stem-diameter per day [1]. Furthermore, stem-diameter enables water stress to be measured continuously in a non-destructive and non-contact manner using a laser displacement sensor that measures the approximate shape of objects based on laser light injected by the sensor. Measuring the stem-diameter with a laser displacement sensor enables us to gather considerable data mechanically necessary for machine learning. Therefore, we adopted stem-diameter as dependent variable. Meanwhile, the installation method and site were determined based on the opinions of experienced farmers.

Although there is a strong relationship between stem-diameter and water stress, it is difficult to use the stem-diameter directly as a dependent variable because it also changes with growth of the plant. Then, we focus on the current difference from the thickest stem-diameter and define the difference in stem-diameter (DSD) as the dependent variable. DSD is expressed as the difference between the maximum stem-diameter (SD) observed thus far and the current stem-diameter (SD_i) as follows:

$$DSD_i = \max(SD_0, \dots, SD_i) - SD_i.$$

The maximum value continuously updates with plant growth. By calculating the decrease from the maximum stem-diameter, the variation due to plant growth is ignored, and only the amount of water stress can be quantified from the stem-diameter.

4. Experiments

4.1. Prototype implementation

We implemented a prototype of the proposed method in a low-stage dense planting cultivation for tomatoes in Shizuoka prefectural agriculture and forestry research institute. For a specific tomato seedling, we installed a small outdoor camera (GoPro HERO 4 Session, Woodman Labs) and a laser displacement sensor (HL-T1010A, Panasonic) for stem-diameter measurement. Moreover, scattered light wireless sensor nodes were installed on and in the plant communities. Our targeted locations were four cultivation beds where twenty-four tomatoes were cultivated in each bed. In our experiment, nursery trees cultivated in the center of each cultivation bed were targeted to predict future water stress. Fig. 8(a), (b), and (c) show an overhead view in one cultivation bed, the layout of measuring equipment for one targeted tomato, and overhead view in entire protected horticulture, respectively. All cameras were attached to the steel pipes of the cultivation beds, and the cameras were installed at locations where the targeted tomato was at the center of images. Paired scattered light sensors installed on and in plant communities can measure temperature, humidity, solar quantity, and plant fertility. In this evaluation, the average

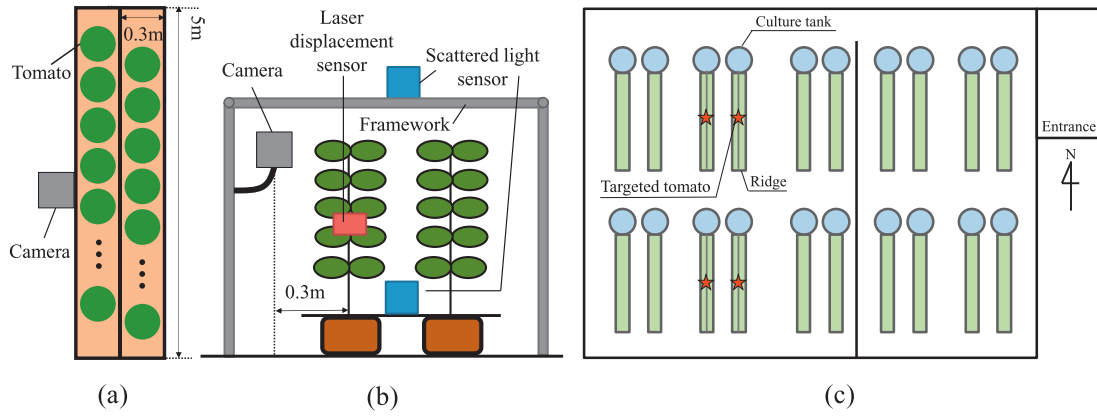


Fig. 8. Experimental environment, (a) overhead view in one cultivation bed, (b) layout of measuring equipment for one targeted tomato, (c) overhead view in entire protected horticulture.

values of the two temperatures and humidities were used as features because these data do not differ significantly between the internal and external plant communities. Meanwhile, scattered light wireless sensor node can measure the two solar quantity types; the upper and inner silicon photodiodes measure direct solar and scattered light, respectively. However direct solar measured by the sensor node installed in plant communities is extremely unstable and must not be used as a feature because the presence or absence of shadows on the sensor changes frequently owing to movement of leaves caused by the wind. Consequently, we used six features: average temperature, average humidity, ratio of paired scattered light, upper direct light, upper scattered light, and inner scattered light. Meanwhile, we replaced the laser displacement sensors at the upper part of the plants after several weeks because stem diameters away from the upper part do not change with growth.

We collected all data during certain growth stages, from the eight-leaf stage for a period of one month. We assumed practical applications for control of the watering systems. It is highly important to control watering precisely during the period to cultivate fruits with high sugar content because plant growth is the fastest during this period. Data was collected at one-minute intervals, and up to 45,000 data were gathered within the period from one cultivation bed.

4.2. Experimental condition

We evaluated the performance of multi-modal SW-SVR using actual agricultural data collected by the prototype. In the evaluation, future DSD was predicted because this knowledge enables farmers to irrigate to the occurrence of severe wilting. According to farmers, tomatoes completely wither within approximately 30 min; then, we predicted the DSD 1 h later with a provisional margin. We performed two evaluations on the detailed performance of multi-modal SW-SVR. In the first evaluation, to investigate DNN using ROAF and environmental data, six features for inputting DNN were compared: ROAF-S, ROAF, POF-S, POF, Org-S, and Org. The details of the comparison are shown in Table 3. The features combining image data and environmental data were inputted to the network architecture as shown in Fig. 7. In contrast, the features with only image data were inputted in another architecture as shown in Fig. 9. The architecture is based on Fig. 7, and the only difference is the omission of the fusion layer. The details of training, validation, and test data are shown in Table 2. The time points all of these data were corrected appropriately and used for evaluation. To evaluate the generalization performance of each feature, cultivation beds of the training data are different to that of the validation and test data. This is because we assume an application that

Table 2

Detail of each data for evaluation.

Name	Image	Environmental data	Network architecture
ROAF-S	ROAF	Used	Fig. 7
ROAF	ROAF	Unused	Fig. 9
POF-S	POF	Used	Fig. 7
POF	POF	Unused	Fig. 9
Org-S	Original	Used	Fig. 7
Org	Original	Unused	Fig. 9

collects data for a certain period and then measures the remaining period using a model tuned with the collected data as with the calibration period of the conventional sensors for agriculture. Therefore, we used the validation data with the same cultivation bed as and different collection period to the test data: the validation data is data collected before 12/Aug./2016, and the test data is the data collected after this in the same area.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2},$$

$$RSE = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad RAE = \frac{\sum_{i=1}^N |y_i - \bar{y}|}{\sum_{i=1}^N |y_i - \bar{y}|},$$

Meanwhile, important parameters are tuned by using random sampling: learning rate, batch size on mini-batch learning, and dropout rate. Finally, the error indicators include mean absolute error (MAE), root mean squared error (RMSE), relative squared error (RSE), and relative absolute error (RAE) as follows: where N is the number of test data, y is the true value, $\bar{y}$ is the average value of all true values, and $\hat{y}$ is the predicted value. All models were tuned based on validation data. The models that have the lowest MAE for validation data were selected as tuned models.

In the second evaluation, we compared the performance of SW-SVR with extracted features by DNN. The comparisons were conventional regression algorithms: decision tree (DT), k-nearest neighbor (k-NN), linear SVR, gradient boosting (GB), random forest (RF), and SVR with a radial basis function kernel. Moreover, to evaluate the superiority of fine-tuning with SW-SVR and the above regression algorithms, DNN was used for comparison. These regression methods are used instead of the output layer in the trained DNN. The error indicators were the same as the first evaluation: MAE, RMSE, RSE, and RAE. Unlike the first evaluation, a grid search is used for parameter tuning.

All implementations for the evaluation were executed in Python. In particular, chainer [29] was used for implementation of DNN, and implementations in scikit-learn [30] were used

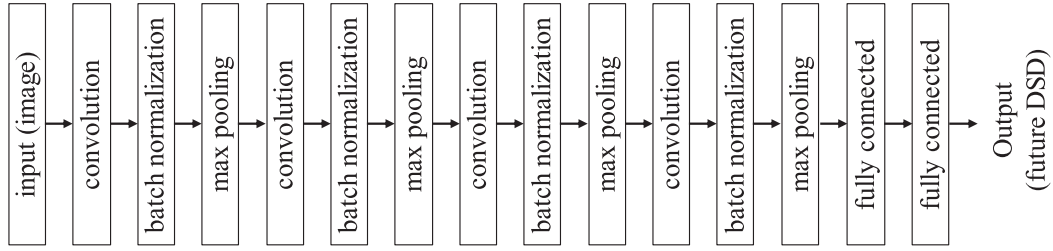


Fig. 9. Network architecture for image-only data.

Table 3

Comparative features for DNN.

Data	Area	Periods	The number of time points
Training data	1, 2, 3	05/Aug./2016–03/Sep./2016	15,778
Validation data	4	05/Aug./2016–12/Aug./2016	1372
Test data	4	12/Aug./2016–03/Sep./2016	4606

Algorithm 1 Sliding window-based support vector regression.

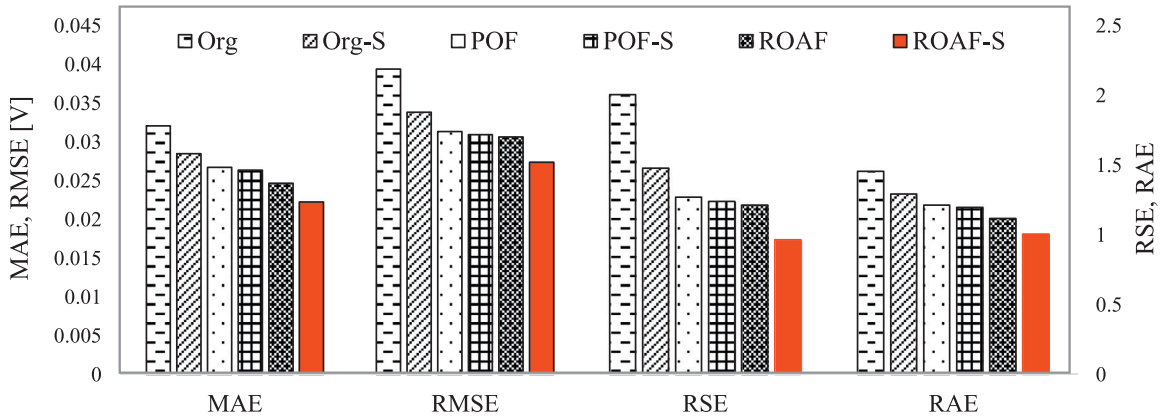
Input:Training data set: $S = \{(\mathbf{x}_1, y_1, \mathbf{x}'_1), \dots, (\mathbf{x}_N, y_N, \mathbf{x}'_N)\}$ where $\mathbf{x}_i \in X$, $y_i \in Y$, $\mathbf{x}'_i \in X'$ Test data: $\mathbf{P}$ Number of models: n Weight parameters: p, q **Preprocessing:**1. Apply normalization to X and X' 2. fit kernel approximation and PLS regression to X and X' 3. $M_i = \|\mathbf{x}_i - \mathbf{x}'_i\|$, $i = 1 \dots N$ 4. $\mathbf{G}_t \leftarrow$ each center of $k\text{menas}(X)$, $t = 1 \dots n$ **For $t = 1$ to n do**1. $D_{ti} = \|\mathbf{G}_t - \mathbf{x}_i\|$, $i = 1 \dots N$ 2. $r_t = \sum_{i=1}^N (w_i M_i) / \sum_{i=1}^N (w_i)$ where $w_i = 1/D_{ti}^p$ 3. $S_t = \{(\mathbf{x}_i, y_i) | D_{ti} < r_t\}$, $i = 1 \dots N$ 4. $H_t(X) \leftarrow$ train $\text{LinearSVR}(S_t)$ **Output:** $H(\mathbf{P}) = \sum_{t=1}^n (w_t H_t(\mathbf{P})) / \sum_{t=1}^n (w_t)$ where $w_t = 1/\|\mathbf{G}_t - \mathbf{P}\|^q$ 

Fig. 10. Error indicators of DNN in test data when using each feature for input of DNN.

for conventional regression algorithms. The code is available at <https://github.com/MinenoLab/MultiModalSWSVR>. This evaluation was performed on a machine with Intel Core i7-5820K Processor, GeForce GTX TITAN X, and 48GB of memory.

4.3. Result and discussion

Fig. 10 shows each error indicator of DNN for the test data when the features are varied. The results show that the prediction performance of ROAF-S is the best, followed in descending order by

that of ROAF, POF-S, POF, Org-S, and Org. In particular, ROAF-S with the best features is able to reduce the prediction error of MAE by approximately 35% compared with Org. Moreover, given that only RSE of ROAF-S was less than 1, the result demonstrates that the prediction of ROAF-S is better than that of the naive model; all prediction values are the average of true values. Meanwhile, the prediction errors are greatly reduced each time the used image data is improved and can be further reduced by adding the environmental data. In particular, the good effect of ROAF is significantly larger

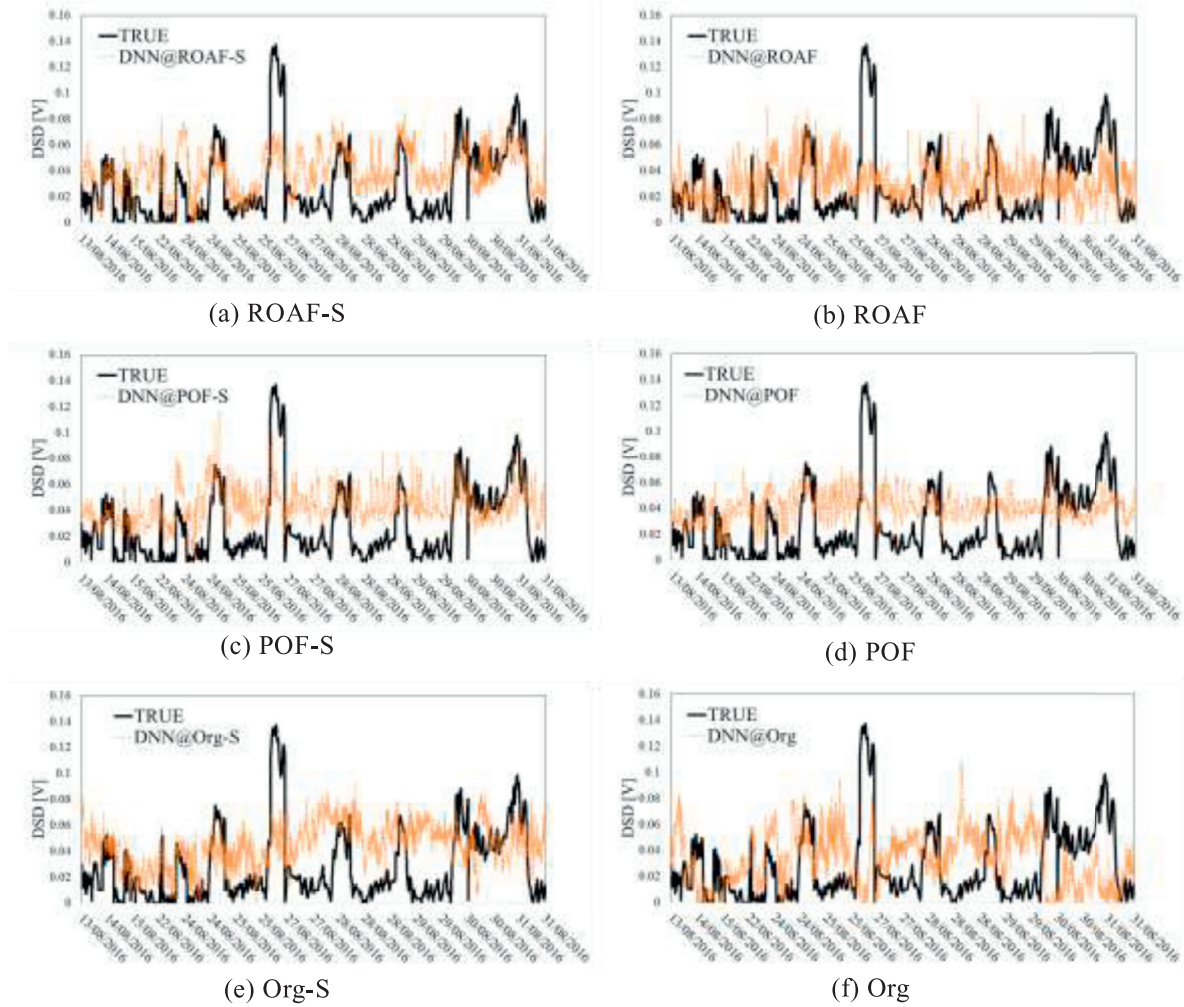


Fig. 11. True values and predicted values of DNN in test data when using each feature for input of DNN.

than that of the environmental data because the prediction error of ROAF is less than that of Org-S.

Fig. 11 shows true values of the test data and predicted values of DNN based on each feature. According to the results, ROAF-S tracks the characteristics of the test data particularly compared to other features. Furthermore, the result clarifies the effects of adding environmental data and ROAF in more detail. Compared with Fig. 11(b) on ROAF and Fig. 11(f) on Org, although all large DSD variation cannot be captured from ROAF, it is possible to calculate the prediction values closer to the true values as a whole because ROAF extracts only important leaves subject to water stress. Specifically, the versatile and essential features are independent of cultivation beds. Meanwhile, the original image data in the same cultivation bed are highly similar to each other because the data is taken continuously from the same location. Consequently, the features depend substantially on the cultivation bed. In this evaluation, the cultivation beds for training and that for validation and testing were divided because practical use is assumed. Therefore, DNN based on the original image data specializes only the cultivation bed for training and predicts considerably different values from true values for test data with a different cultivation bed. On the other hand, the model based on ROAF involves high generalization performance for test data of different cultivation beds. Next, compared with Fig. 11(a), (c), and (e) on using environmental data and Fig. 11(b), (d), and (f) on not using that, the addition of environmental data allows the possibility of DNN with ROAF to

capture large variations of DSD more correctly. The result demonstrates that ROAF, which is versatile and an essential image features related to water stress, also increases the effect including environmental data.

Next, Fig. 12 shows each error indicator when SW-SVR and the comparative models predict future DSD using the extracted features by DNN based on ROAF-S. The result demonstrates that the prediction performance of SW-SVR is best in the models. In particular, SW-SVR is able to reduce the prediction error of MAE by approximately 20% compared with that of DNN. Moreover, although other prediction errors are worse compared to DNN, SW-SVR is the only model that can use the extracted features based on ROAF-S effectively. Fig. 13 shows the true values and predicted values of DNN based on ROAF-S, SW-SVR, and SVR, which is the best model in the comparison. SW-SVR outputs the predicted values close to the true values compared to SVR, and the prediction is highly stable. In particular, the test data predicted largely in DNN can be predicted accurately in SW-SVR, and it is possible to capture the characteristic variation of DSD correctly. In contrast, the predicted values of SVR closely approximate that of DNN, to which substantial noise is added; the predicted values of SVR was more unstable than that of DNN. Consequently, when using the conventional regression algorithms, it is difficult to make better relationships between the dependent variable and independent variables extracted by DNN than the output layer of DNN. Meanwhile, the reason for the reduction in the prediction error in SW-SVR is that the charac-

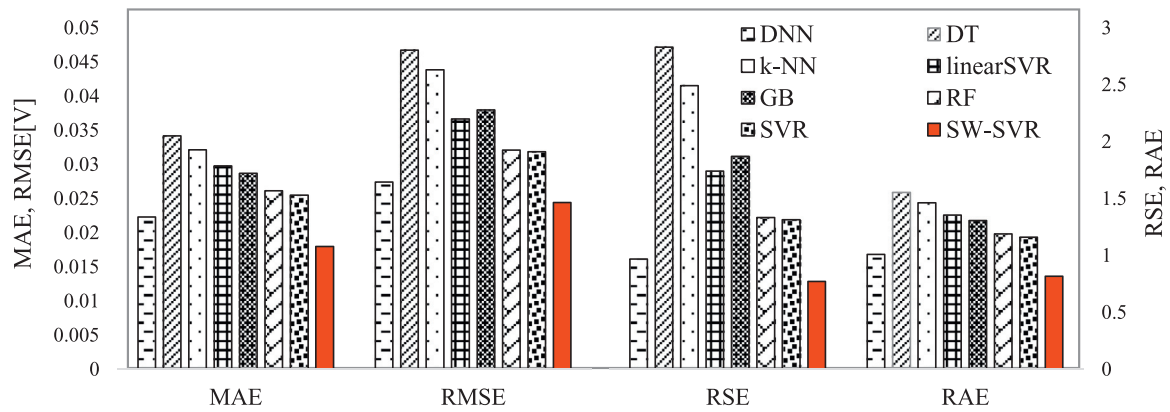


Fig. 12. Error indicators of each model in test data when using ROAF-S for input of DNN.

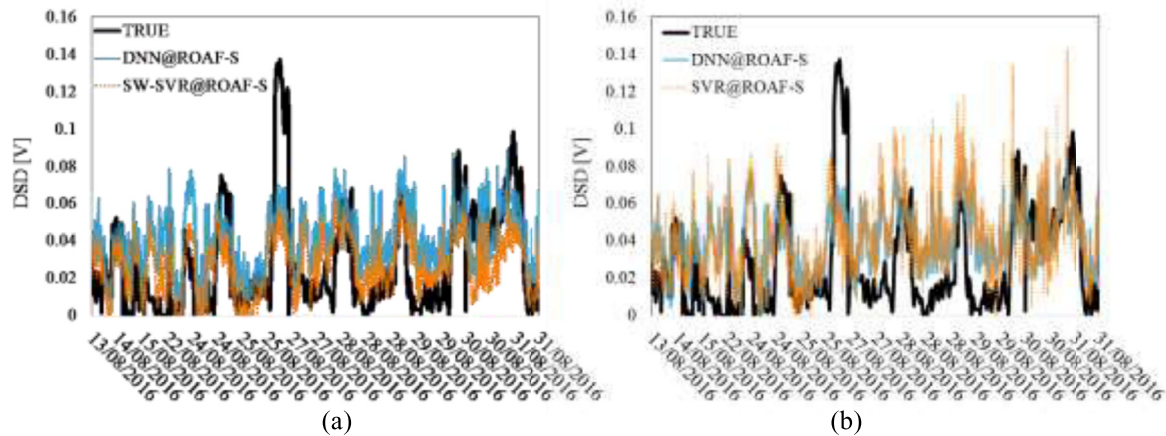


Fig. 13. True values and predicted values when using ROAF-S for input of DNN, (a) compared with DNN and SW-SVR, (b) compared with DNN and SVR that is the best model in all comparisons.

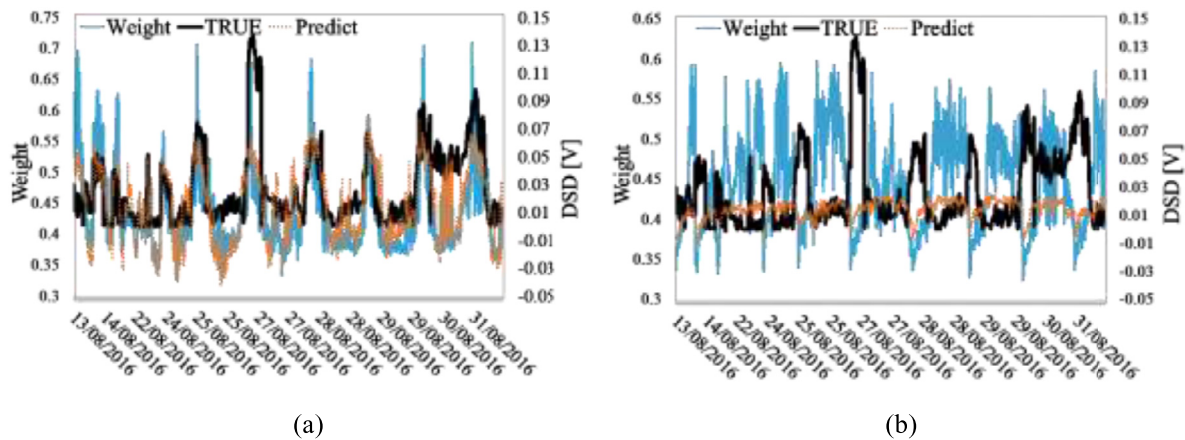


Fig. 14. True values, weights, and predicted values for each specialized model in test data, (a) the specialized model for high water stress, (b) the specialized model for low water stress.

teristic variation with time in untrained test data can only be captured with the data extraction of D-SDC and dynamic weighting in SW-SVR.

Fig. 14 indicates the weights of each specialized model, true values of test data, and predicted values of each specialized model when focusing on specialized models for two situations: high water stress and low water stress. Note that whether the models are specialized in which situations is decided from an average of the dependent variable of the training data used for building each spe-

cialized model. That of the specialized model for high water stress shown in Fig. 14(a) is 0.043, and that of the specialized model for low water stress shown in Fig. 14(b) is 0.014. Fig. 14(a) shows that the specialized model for high water stress can predict data with large DSD relatively accurately, at the same time, weights for integration when predicting the data is increased. On the other hand, data with low water stress outside the specialization is not predicted correctly, but the weights when predicting the data is reduced. Fig. 14(b) further shows that the specialized model for low

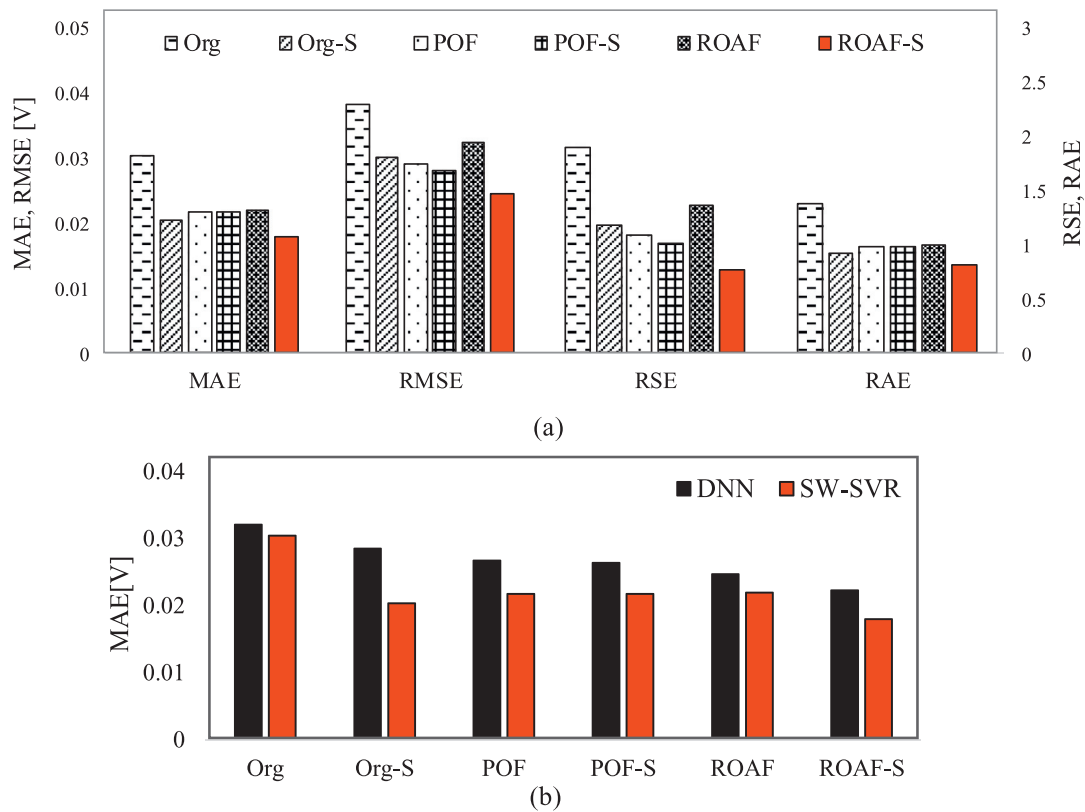


Fig. 15. Error indicators of SW-SVR in test data when using extracted features by DNN, (a) error indicators of SW-SVR when using each feature for input of DNN, (b) MAE of SW-SVR and DNN when using each feature for input of DNN.

water stress can predict only data with low water stress more accurately, and the weights increase when only predicting the data with low water stress. These results demonstrate the effectiveness of SW-SVR from two viewpoints. First, specialized models built by extracted training data by D-SDC can predict only specialized situations accurately. Second, dynamic weighting is strongly affected by each specialized model only when the models can accurately predict the test data with time-dependent characteristics. Specifically, the specialized models for various situations have a great relationship in which mutual strength and weaknesses are interpolated with each other. Based on above advantage in SW-SVR, it is believed that it is the only model that can make full use of the multi-modal features extracted by DNN effectively. The first evaluation mentioned above describes the effectiveness of extracting multi-modal features using DNN, and this second evaluation mentions that SW-SVR can build a more effective model based on the extracted multi-modal features by novel approaches not characteristic of other models. From these discussions, it is meaningful to distinguish between DNN for extracting multi-modal features and SW-SVR for future prediction.

Finally, Fig. 15 shows each error indicator of SW-SVR for varied features in DNN. Fig. 15(a) shows that ROAF-S is the best feature for predicting future water stress even when the processing of the output layer in DNN is replaced with SW-SVR. In particular, the features based on ROAF-S reduced the prediction error of MAE by approximately 41% compared with those based on Org in SW-SVR. Meanwhile, compared with DNN and SW-SVR as shown in Fig. 15(b), SW-SVR demonstrates better prediction performance than the output layer of DNN for all features. Although the best feature for SW-SVR is ROAF-S, the differences between ROAF-S and other features in SW-SVR are smaller than that of DNN. Scatter plots of SW-SVR predicted values with respect to the true values of DSD using each feature for input of DNN are shown in Fig. 16.

The results indicate that the ROAF-S has the highest positive correlation coefficients of all extracted features. Although, it is believed that SW-SVR can build a model effectively for various features extracted by DNN, further studies are required to reveal the detailed performance of multi-modal SW-SVR.

5. Conclusion

We proposed a novel plant water stress prediction method using multi-modal SW-SVR to reproduce farmers' cultivation precisely. Given that plant image data and environmental data complement each other's insufficient information, the complicated water stress is expressed multilaterally. In multi-modal SW-SVR, DNN is applied to feature extraction in conventional SW-SVR to predict water stress accurately based on the environmental and plant image data. DNN extracts multi-modal features related to water stress using the proposed image feature, ROAF, which expresses image data of leaves only using water stress based on plant wilting motion. ROAF reproduces the same perspective on the original image data as farmers, and DNN can be learned with the viewpoint in complicated plant image data and environmental data. We evaluated the proposed multi-modal SW-SVR with ROAF using actual agricultural data. The experimental results demonstrated that the model built from the multi-modal features combining ROAF and environmental data, surpasses that based only on original image data significantly in terms of prediction error. Moreover, the proposed ROAF enables SW-SVR to predict future water stress more accurately than DNN and existing regression algorithms.

In future work, the introduction of motion information must be considered. Although ROAF focuses on leaves with large movements owing to water stress, the information regarding magnitude and angle obtained in optical flow have not been used directly. However, problems exist in applying the optical flow directly to

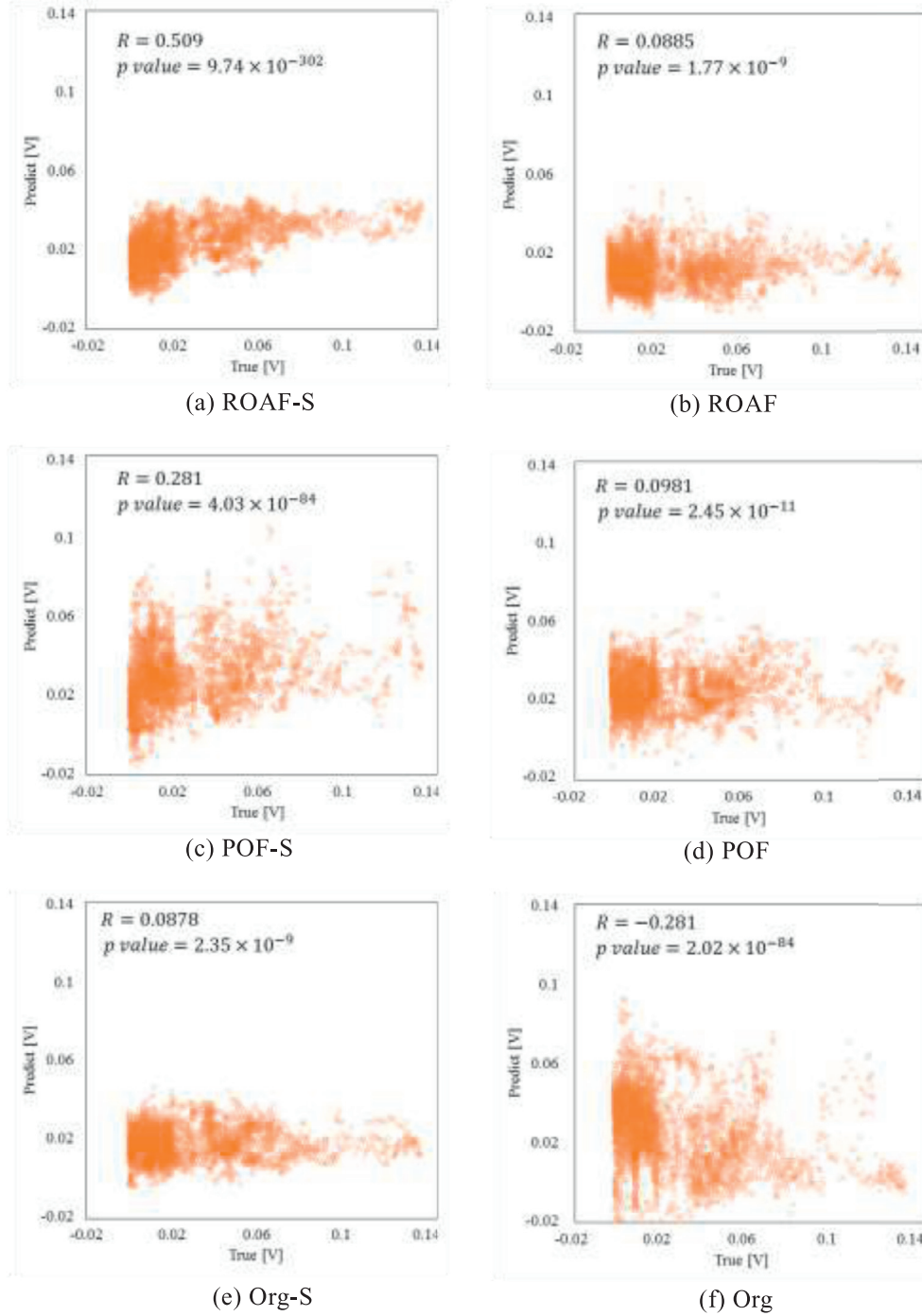


Fig. 16. Scatter plots of predicted values with SW-SVR with respect to true values when using each feature for input of DNN.

DNN because the former is highly dependent on the camera location and shoot date, resulting in different scales of optical flow for each image data. The scale difference increases considerably if both the cultivation beds and the shoot date differ. Therefore, normalization of the optical flow must be taken into consideration to eliminate the scale. Next, we built a model with validation data and test data using the same cultivation bed. This model needs to be calibrated for each field as required when in use. Therefore, we will train the model using new training, validation, and test data in a completely separate cultivation bed. Meanwhile, as another approach to extract features related to water stress more faith-

fully in DNN, we consider using a recurrent neural network (RNN), such as long short-term memory (LSTM), which can consider the temporal relationship of a specific period. We will aim to capture the short-term and long-term time relationship by further devising an image feature substitute for ROAF and DNN. Moreover, multi-modal SW-SVR is a generic method rather than a specific method for a specific application such as water stress prediction. Therefore, we would evaluate the performance using a more general dataset, such as meteorological data and moving objects.

Acknowledgements

This work was supported by JST, PRESTO Grant Number JPMJPR1505, Japan. We greatly appreciate Mr. Oishi, Mr. Imahara and Mr. Maejima, Shizuoka prefectural research institute of agriculture and forestry.

References

- [1] X. Wang, Z. Meng, X. Chang, Z. Deng, Y. Li, M. Lv, Determination of a suitable indicator of tomato water content based on stem diameter variation, *Scientia Horticulturae* 215 (2017) 142–148. <http://doi.org/10.1016/j.scienta.2016.11.053>.
- [2] M. Sano, Y. Nakagawa, T. Sugimoto, T. Shirakawa, K. Yamagishi, T. Sugihara, M. Ohaba, S. Shibusawa, Estimation of water stress of plant by vibration measurement of leaf using acoustic radiation force, *Acoust. Sci. Technol.* 36 (3) (2015) 248–253. <http://doi.org/10.1250/ast.36.248>.
- [3] J.A. Sánchez-Molina, F. Rodríguez, J.L. Guzmán, J.A. Ramírez-Arias, Water content virtual sensor for tomatoes in coconut coir substrate for irrigation control design, *Agr. Water Manage.* 151 (2015) 114–125. <http://doi.org/10.1016/j.agwat.2014.09.013>.
- [4] K. Takayama, H. Nishina, Early detection of water stress in tomato plants based on projected plant area, *Environ. Control Biol.* 45 (4) (2007) 241–249. <http://doi.org/10.2525/ecb.45.241>.
- [5] C. Patané, S.L. Cosentino, Effects of soil water deficit on yield and quality of processing tomato under a Mediterranean climate, *Agr. Water Manage.* 97 (1) (2010) 131–138. <http://doi.org/10.1016/j.agwat.2009.08.021>.
- [6] C. Duarte-Galvan, R.D.J. Romero-Troncoso, I. Torres-Pacheco, R.G. Guevara-Gonzalez, A.A. Fernandez-Jaramillo, L.M. Contreras-Medina, R.V. Carrillo-Serrano, J.R. Millan-Almaraz, FPGA-based smart sensor for drought stress detection in tomato plants using novel physiological variables and discrete wavelet transform, *Sensors* 14 (10) (2014) 18650–18669. <http://doi.org/10.3390/s141018650>.
- [7] Z.Q. Yuan, K.L. Yu, B.X. Wang, W.Y. Zhang, X.L. Zhang, K.H. Siddique, K. Stefanova, N.C. Tuner, F.M. Li, Cutting improves the productivity of lucerne-rich stands used in the revegetation of degraded arable land in a semi-arid environment, *Sci. Rep.* 5 (2015). <http://doi.org/10.1038/srep12130>.
- [8] R.C. Yang, Towards understanding and use of mixed-model analysis of agricultural experiments, *Can. J. Plant Sci.* 90 (5) (2010) 605–627. <http://doi.org/10.4141/CJPS10049>.
- [9] A. Singh, B. Ganapathysubramanian, A.K. Singh, S. Sarkar, Machine learning for high-throughput stress phenotyping in plants, *Trends Plant Sci.* 21 (2) (2016) 110–124. <http://doi.org/10.1016/j.tplants.2015.10.015>.
- [10] Y. Suzuki, Y. Kaneda, H. Mineno, Analysis of support vector regression model for micrometeorological data prediction, *Comput. Sci. Inf. Technol.* 3 (2) (2015) 37–48. <http://doi.org/10.13189/csit.2015.030202>.
- [11] Namin, S.T., Esmailzadeh, M., Najafi, M., Brown, T.B., & Borevitz, J.O. (2017). Deep phenotyping: deep learning for temporal phenotype/genotype classification. *bioRxiv*, 134205. <https://doi.org/10.1101/134205>.
- [12] Y. Kaneda, H. Mineno, Sliding window-based support vector regression for predicting micrometeorological data, *Expert Syst. Appl.* 59 (2016) 217–225. <http://doi.org/10.1016/j.eswa.2016.04.012>.
- [13] D.H. Park, J.W. Park, Wireless sensor network-based greenhouse environment monitoring and automatic control system for dew condensation prevention, *Sensors* 11 (4) (2011) 3640–3651. <http://doi.org/10.3390/s110403640>.
- [14] J. Boyer, Leaf water potentials measured with a pressure chamber, *Plant Physiol.* 42 (1) (1967) 133–137.
- [15] J. Macqueen, Some methods for classification and analysis of multivariate observations, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* 1 (1967) 281–297. <http://doi.org/citeulike-article-id:6083430>.
- [16] A. Rahimi, B. Recht, Random features for large-scale kernel machines, *Adv. Neural Inf. Process. Syst.* 3 (4) (2007) 5. <http://doi.org/10.1.1.145.8736>.
- [17] M. Tenenhaus, V.E. Vinzi, Y.M. Chatelin, C. Lauro, PLS path modeling, *Comput. Statist. Data Anal.* 48 (1) (2005) 159–205. <http://doi.org/10.1016/j.csda.2004.03.005>.
- [18] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, *Adv. Neural Inf. Process. Syst.* 19 (2007) 153. <http://doi.org/citeulike-article-id:4640046>.
- [19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* 15 (1) (2014) 1929–1958. <http://doi.org/10.1214/12-AOS1000>.
- [20] S. Ioffe, C. Szegedy, Batch Normalization: Accelerating deep network training by reducing internal covariate shift, 2015. *arXiv:1502.03167* <http://doi.org/10.1007/s13398-014-0173-2>.
- [21] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034. <http://doi.org/10.1109/ICCV.2015.123>.
- [22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9. <http://doi.org/10.1109/CVPR.2015.7298594>.
- [23] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. <http://doi.org/10.3389/fpsyg.2013.00124>.
- [24] Y. Kim, H. Lee, E.M. Provost, Deep learning for robust feature generation in audiovisual emotion recognition, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 3687–3691. <http://doi.org/10.1109/ICASSP.2013.6638346>.
- [25] H. Ibayashi, Y. Kaneda, J. Imahara, N. Oishi, M. Kuroda, H. Mineno, A reliable wireless control system for tomato hydroponics, *Sensors* 16 (2016).
- [26] P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid, DeepFlow: large displacement optical flow with deep matching, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1385–1392. <http://doi.org/10.1109/ICCV.2013.175>.
- [27] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2–9. <http://doi.org/10.1109/CVPR.2009.5206848>.
- [28] C. Feichtenhofer, A. Pinz, A. Zisserman, Convolutional two-stream network fusion for video action recognition, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2016, pp. 1933–1941. <http://doi.org/10.1109/CVPR.2016.213>.
- [29] S. Tokui, K. Oono, S. Hido, J. Clayton, Chainer: a next-generation open source framework for deep learning, in: *Proceedings of Workshop on Machine Learning Systems in the Twenty-Ninth Annual Conference on Neural Information Processing Systems*, 2015, pp. 1–6.
- [30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830. <http://doi.org/10.1007/s13398-014-0173-2>.