



OPEN

UAV-based individual Chinese cabbage weight prediction using multi-temporal data

Andrés Aguilar-Ariza¹, Masanori Ishii², Toshio Miyazaki³, Aika Saito³, Hlaing Phyo Khaing³, Hnin Wint Phoo³, Tomohiro Kondo³, Toru Fujiwara¹, Wei Guo² & Takehiro Kamiya¹✉

The use of unmanned aerial vehicles (UAVs) has facilitated crop canopy monitoring, enabling yield prediction by integrating regression models. However, the application of UAV-based data to individual-level harvest weight prediction is limited by the effectiveness of obtaining individual features. In this study, we propose a method that automatically detects and extracts multitemporal individual plant features derived from UAV-based data to predict harvest weight. We acquired data from an experimental field sown with 1196 Chinese cabbage plants, using two cameras (RGB and multi-spectral) mounted on UAVs. First, we used three RGB orthomosaic images and an object detection algorithm to detect more than 95% of the individual plants. Next, we used feature selection methods and five different multi-temporal resolutions to predict individual plant weights, achieving a coefficient of determination (R^2) of 0.86 and a root mean square error (RMSE) of 436 g/plant. Furthermore, we achieved predictions with an R^2 greater than 0.72 and an RMSE less than 560 g/plant up to 53 days prior to harvest. These results demonstrate the feasibility of accurately predicting individual Chinese cabbage harvest weight using UAV-based data and the efficacy of utilizing multi-temporal features to predict plant weight more than one month prior to harvest.

Vegetable production is often hindered by adverse environmental conditions, such as climate and weather variability^{1,2}. To overcome these challenges and ensure stable production, genetic improvements through breeding³ and anticipatory crop-management corrections¹ are needed. The implementation of these solutions requires the collection of phenotypic data throughout the growth cycle. However, this process is typically performed manually, which can be both costly and time-consuming, especially for crops that require individual plant characterization⁴.

Remote sensing data have become a popular alternative for crop phenotyping because they can provide information on a large scale in a non-expensive and non-destructive manner⁵. The rapid development of sensors and platforms has boosted remote sensing surveys using devices such as unmanned aerial vehicles (UAVs). Owing to their versatility in capturing data at different spatial and temporal resolutions^{6,7}, features obtained from sensors such as visible-light (RGB) cameras, multi-spectral (MS) cameras, light detection and ranging systems (LiDAR), and thermal infrared imagery have been used for crop phenotyping^{3,7}. Furthermore, in combination with image analysis techniques, additional features derived from the sensors are computed to represent the crop morphological and physiological attributes^{4,8,9}. For example, techniques such as structure-from-motion (SfM), used for 3D point cloud reconstruction, and vegetation indices (VIs) computed from spectral reflectance imagery provide important insights into plant structure and nutritional conditions^{10,11}. These multi-sourcing features, in combination with statistical and machine learning regression methods, have provided promising results in biomass prediction for crops such as coffee¹², potatoes¹³, fava beans¹⁴, alfalfa⁹, soybeans¹⁰, wheat¹⁵, and cotton^{16,17}.

Despite the advances in yield prediction using crop phenotyping with UAV-based data, these data are typically obtained at the canopy level rather than at the individual plant level^{9,18}. However, this approach hinders the phenotyping of individually sold vegetables, such as Chinese cabbage, where pricing is determined on the basis of weight per individual plant rather than weight per unit area⁴. Although Chinese cabbage is an economically important crop in East Asia¹⁹, few studies have investigated the use of UAV-based phenotyping of individual plants. For example, Kang et al.²⁰ used a multi-spectral camera to acquire data at a single time point to predict the weight of Chinese cabbage. Similarly, Kim et al.²¹ predicted multiple morphological attributes including fresh

¹Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. ²Institute for Sustainable Agro-Ecosystem Services, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1, Midoricho, Nishitokyo-shi, Tokyo 188-0002, Japan. ³Nippon Norin Seed Co., 6-6-5 Takinogawa, Kita-ku, Tokyo 114-0023, Japan. ✉email: akamiyat@g.ecc.u-tokyo.ac.jp

weight by extracting UAV-based data using bounding boxes. Although both studies predicted Chinese cabbage weight using either multi-spectral or RGB imagery, individual plant data were extracted by applying pixel-based segmentation to areas that were manually located to represent the plant location.

To efficiently obtain individual plant multi-temporal UAV-based data, we developed an approach to automatically detect individual Chinese cabbages in the field using an object detection algorithm known as YOLOv5²², and predicted individual plant weights using machine learning models. Furthermore, we evaluated multi-temporal features to predict Chinese cabbage weight several days prior to harvest. Our approach demonstrates the feasibility of predicting individual Chinese cabbage weights using UAV-based data up to 53 days before harvest with an RMSE of 560 g/plant and $R^2 = 0.72$.

Results

UAV flights and weight measurement

The Chinese cabbage growth period in the field was 104 days, during which the UAV collected data from 26 different time points (TPs) using two cameras (RGB and MS) (Table S1). During each flight, 67 and 133 images were captured using the RGB and MS cameras, respectively.

In total 1196 plants that were planted in the field, we measured the weight of 872 plants; the remaining plants were not included, mainly because of workforce limitations. The average plant weight measured was 2847 ± 940 g/plant. Weight variability can be attributed to differences in the genotype of the F2 population.

Individual plant detection

To save time in detecting individual plants, we applied an object-detection algorithm instead of manually drawing a bounding box for each plant. We used an object-detection algorithm based on convolutional neural networks (YOLO) to automatically locate individual Chinese cabbage plants. To evaluate the accuracy of the detection algorithm, the field was divided into two groups: training ($n = 589$) and testing ($n = 547$). Once the model was trained, it was applied to the test dataset. The model detected 469 (85.7%), 528 (96.5%), and 539 (98.5%) plants on September 29, October 4, and October 6, respectively (Fig. S1). The lowest accuracy (85.7%) was obtained for the first date (12 DAT).

Then, we merged the bounding boxes (see the “Materials and methods” section). This enabled us to obtain the final individual plant bounding box. A total of 1136 individual plants were detected, representing 95% of the total number of plants (Fig. 1A). Plants that were not detected had smaller leaf diameters than other plants (Fig. 1A; Fig. S1). Finally, to extract information from the RGB and MS orthomosaic images (Fig. 1B), we transformed the bounding box coordinate system from the image spatial reference to the projected coordinate system (UTM zone 54N), in which the UAV captured the data.

Search for the most relevant features for individual plant weight prediction

In total, 18 features were computed for each identified plant across 26 TPs. These features were grouped into three levels: the first level included eight features derived from the RGB and MS orthomosaic imagery, the second level included seven VIs, and the third level encompassed three features derived from the 3D point cloud data. We reduced the number of features to reduce multicollinearity. Three feature selection methods were used: two based

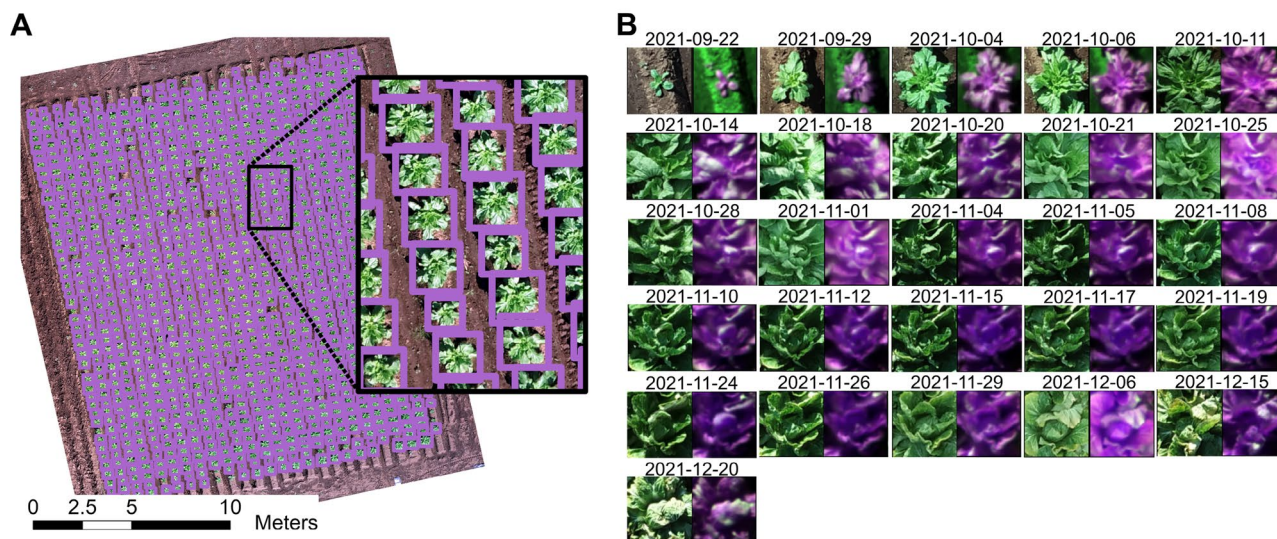


Figure 1. Individual plant detection. (A) The final bounding boxes (purple) were obtained by merging the bounding boxes for three TPs (September 29, October 4, and October 6, 2021). The black box is an enlarged area as an example of plants detected by the model. Figure prepared using ArcGIS Pro version 3.0²⁴ (B) An example of the multi-temporal RGB and MS imagery extracted for an individual plant. On the left is the true-color RGB image, while on the right is the false-color MS image (NIR—red-edge—red). The individual plant RGB and MS imagery were obtained with the Matplotlib package in Python version 3.9^{25,26}.

on wrapper methods [exhaustive search and sequential feature selection (SFS)] and one based on an embedded method [random forest (RF)]²⁷. The most recent TP data (December 20, 2021) were used to train each filtering method. Out of 872 plants weighed, 33 plants were not detected by YOLOv5 and therefore 839 plants were used for the weight prediction. We split them into two datasets: a training dataset (n = 756) and a test dataset (n = 83).

For the first level, the exhaustive search and SFS methods yielded similar accuracies ($R^2 = 0.16$) to those using more features. For example, the incorporation of four features resulted in the highest R^2 of 0.17 (Fig. 2A). As no

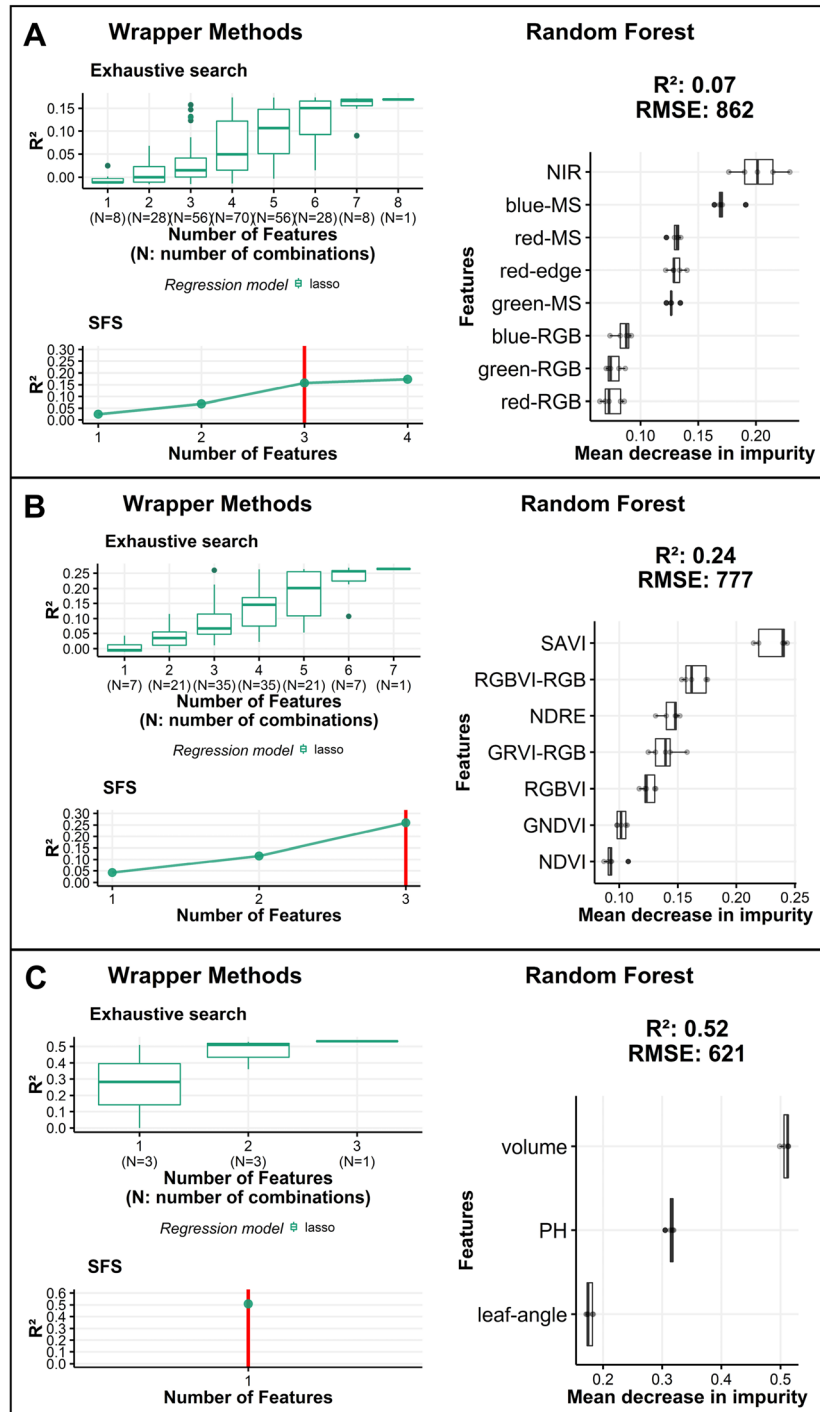


Figure 2. Wrapper [exhaustive research and sequential feature selection (SFS)] and embedded [random forest (RF)] method results for three levels of information. (A) First level: images taken directly with RGB and MS cameras. (B) Second level: vegetation indices (VIs). (C) Third level: features derived from the 3D point cloud data. The red line in the SFS plots represents the number of selected features. Box plot and bar plot figures were created with the ggplot2 package in R version 4.1.1.^{28,29}

significant improvement incurred by including more than three features, we selected features that occupied the top three positions in the RF importance index [Near Infrared (NIR), blue-MS, and red-MS]. For the second level, when the wrapper method was applied to the seven VIs, the models trained with three VIs yielded the highest accuracy ($R^2 = 0.26$) (Fig. 2B). Although the top three features ranked by the RF were SAVI, RGBVI-RGB, and NDRE, both wrapper methods yielded the highest accuracy using SAVI, RGBVI-RGB, and GRVI-RGB. Therefore, these three features were selected for this study. Finally, for the third level, the exhaustive search indicated that by using one variable (volume), the models achieved an R^2 of 0.51, and by including all three variables, the models achieved an R^2 of 0.53 (Fig. 2C). Similarly, the SFS and RF indicated that volume was the most relevant feature for predicting weight. Therefore, volume was selected as the feature for the third level. In summary, for individual plant weight prediction, seven features were selected: three from the first level (red-MS, blue-MS, and NIR), three from the second level (SAVI, RGB-RGB, and GRVI-RGB), and one from the third level (volume).

Assessing the selected features with five regression models

To evaluate the feature selection results, we trained five regression models: partial least squares (PLS), lasso, ridge, support vector machine (SVM-linear), and RF. We used the last TP data to train the models with the five-fold cross-validation datasets, which were the same as those used for feature selection. For each level of information, we trained the regression models using all the features (Fig. 3) or selected features (Fig. 3). For example, at the first level, there were eight features before selection (Table S2) and three features (red-MS, blue-MS, and NIR) after selection. In addition, we trained the regression models using all 18 features and the seven selected features (all levels in Fig. 3).

The best model performance was achieved when all levels of information were combined ($R^2 = 0.64$; RMSE = 541 g/plant) (Fig. 3. Before feature selection using Ridge), whereas the worst prediction was obtained by the models that only used features from the first level (Fig. 3). Although the model with the best accuracy was the one trained with all 18 features as input, the models trained with the seven selected features achieved similar accuracies ($R^2 = 0.62$; RMSE = 556 g/plant) (Fig. 3. After feature selection using PLS). These results indicate that,

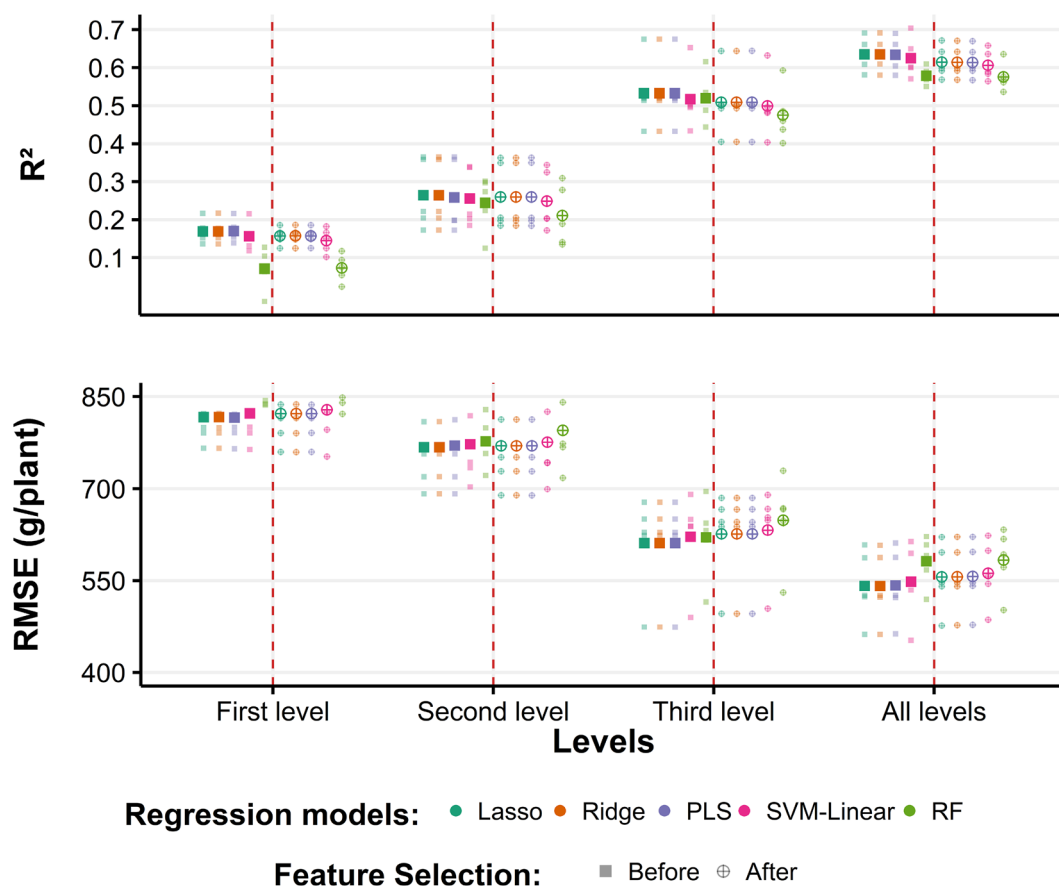


Figure 3. Comparison of the regression model predictions, before and after applying the feature selection methods on three levels of information (RGB and MS image layers, vegetation indices, and features derived from the 3D point cloud data) and combining all levels. The small dots represent each of the five-fold cross-validation results. The red dotted line separates the evaluation metrics before (left) and after (right) feature selection. The colors indicate each regression model [lasso, ridge, PLS, SVM-linear, random forest (RF)]. The small and large points represent the result of five-fold cross-validation and the mean value of the cross-validation results, respectively. Dot plots were created with the ggplot2 package in R version 4.1.1.^{28,29}

regardless of the reduction in the number of features, the models did not lose their prediction capability. The RF models exhibited the lowest performance at all levels (represented in green in Fig. 3).

Weight prediction using multi-temporal data

The use of multi-temporal data for weight prediction was evaluated using the following two questions: (1) Does weight prediction improve when multi-temporal data are involved? (2) What is the optimal UAV multi-temporal resolution for weight estimation? To answer these questions, we evaluated the values predicted by the regression models using five-fold cross-validation and a test dataset. The models were trained using the seven selected features computed for the 26 TPs.

For the first question, we separately trained the regression models with features extracted from two temporal schemes: single- and multi-temporal points. The single time-point scheme was computed for a specific TP, whereas the multi-temporal point scheme used not only specific TP data but also sequencing data from the first TP (12 DAT). We used five-fold cross-validation and test datasets to compare the weight predictions using both schemes. The best average cross-validation accuracy was obtained by the SVM-linear model using multi-temporal data between 12 and 96 DAT ($R^2 = 0.73$; RMSE = 464 g/plant); for the single time-point scheme, the best cross-validation results were obtained by the PLS model using data at 77 DAT ($R^2 = 0.66$; RMSE = 525 g/plant) (Fig. 4). For the test dataset, the SVM-linear model trained with a multi-temporal scheme yielded the best accuracy using data up to 96 DAT ($R^2 = 0.83$; RMSE = 436 g/plant), whereas the lasso model achieved the best prediction accuracy for the single time-point scheme using data at 70 DAT ($R^2 = 0.72$; RMSE = 557 g/plant) (Fig. 4). These results show that when multi-temporal data were used to train the models, the accuracy of the weight prediction improved.

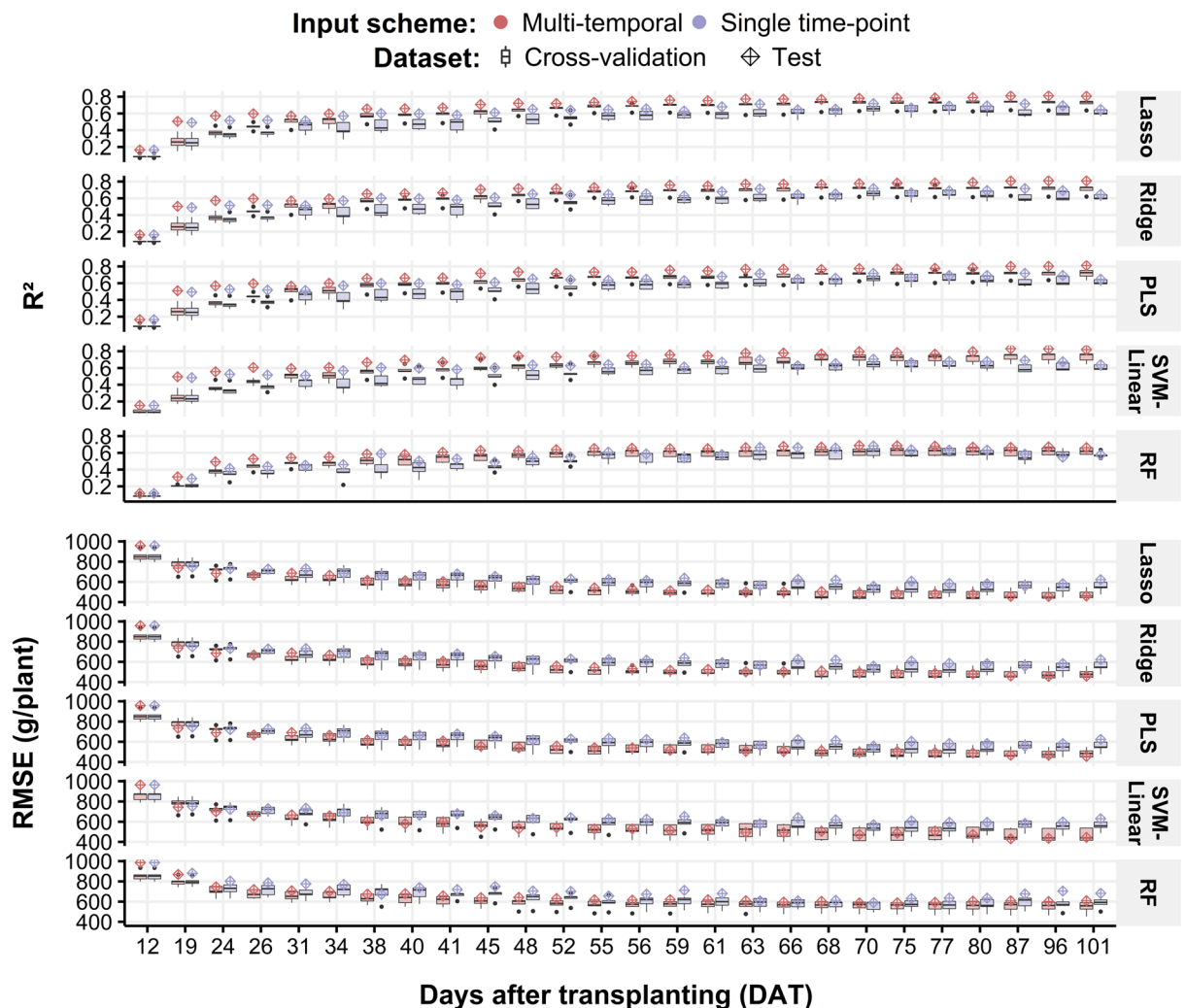


Figure 4. R^2 (top) and RMSE (bottom) of the five-fold cross-validation and test results for each regression model [lasso, ridge, PLS, SVM-linear, and random forest (RF)]. The results were compared for two input schemes: multi-temporal (red) and single-time-point (purple) data. The boxplots show a five-fold cross-validation results distribution, whereas the diamond points indicate the test dataset results. Box plots were created with the ggplot2 package in R version 4.1.1.^{28,29}

For the second question, we used five multi-temporal resolutions (Fig. 5A). Four resolutions include a reduced number of TPs and one resolution was represented by all 26 TPs (“All TPs” in Fig. 5A). To reduce the number of TPs, three multi-temporal resolutions were set based on fixed time intervals (7-, 15-, and 30-day intervals) and one resolution (growth pattern) was set based on the Chinese cabbage growth cycle (12, 34, 56, 80, and 101 DAT; Fig. S2). As a result, at each multi-temporal resolution, 26, 14, 7, 4, and 5 TPs were used for all TPs, 7-day, 15-day, 30-day, and growth patterns, respectively (Fig. 5A). To compare the prediction accuracy obtained among the five multi-temporal resolutions, we used the fivefold cross-validation and test datasets. The best averaged cross-validation accuracy prediction was obtained by the 7-day multi-temporal resolution using data from 14 TPs ($R^2=0.74$; RMSE = 457 g/plant), and a similar accuracy was obtained when the models were trained using all 26 TPs ($R^2=0.73$; RMSE = 466 g/plant) and seven TPs ($R^2=0.72$; RMSE = 473 g/plant) (Fig. 5B). For the test dataset, the best accuracy was obtained by the 7-day multi-temporal resolution ($R^2=0.82$; RMSE = 444 g/plant), followed by all 26 TPs ($R^2=0.81$; RMSE = 444 g/plant) (Fig. 5B) and 15-day (7 TPs) resolutions ($R^2=0.79$; RMSE = 479 g/plant). These results indicate that regardless of the reduction in the number of TPs from 26 to 14 and 7, the regression models can still achieve reliable predictions.

Weight prediction prior to harvest

To assess the ability to predict Chinese cabbage weight before harvest, we trained regression models using multiple sequences of TP data. For example, to predict Chinese cabbage weight using data up to 30 days prior to harvest (DPH), we used 20, 9, 4, 3, and 3 TP data points for each of the five multi-temporal resolutions (all TPs, 7-day, 15-day, 30-day, and growth patterns). The weight prediction was performed using the test dataset. The best prediction accuracy ($R^2=0.83$; RMSE = 436 g/plant) was obtained by the SVM-linear model using the sequence TP data between the transplantation date (12 DAT) and 5 DPH (96 DAT) for all TP multi-temporal resolutions (Fig. 6, red lines). For the other multi-temporal resolutions, the 7-day resolution achieved the highest prediction accuracy ($R^2=0.83$; RMSE = 453 g/plant) when the model used 12 TP-sequenced data up to 14 DPH (87 DAT) (Fig. 6, blue lines). In addition, the 15-day interval models yielded an $R^2=0.81$ and RMSE = 455 g/plant using 6 TP (87 DAT, 14 DPH) (Fig. 6, green lines). Furthermore, the models trained with sequenced data up to 53 DPH yielded R^2 greater than 0.72 and RMSE less than 560 g/plant (Fig. 6, in PLS). This was achieved using 7-day and all TP multi-temporal resolutions. At 15 days, the earliest date to obtain predictions with R^2 greater than 0.72 ($R^2=0.77$; RMSE = 505 g/plant) was 31 DPH (Fig. 6).

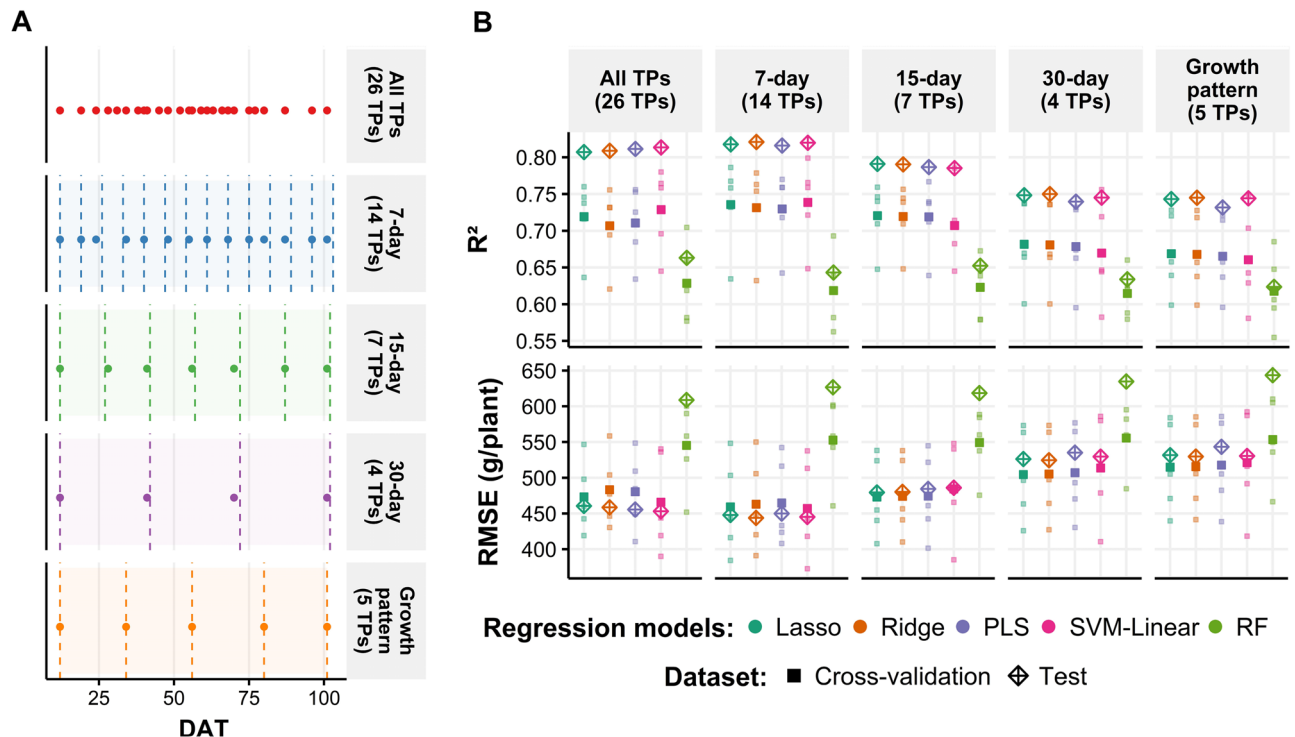


Figure 5. Individual plant weight prediction using 5 different multi-temporal resolutions. (A) The panel shows the selected TPs in each multi-temporal resolution. (B) The R^2 and RMSE five-fold cross-validation and test results for 5 different multi-temporal resolutions. The colors indicate each regression model [lasso, ridge, PLS, SVM-linear, random forest (RF)]. The small and large square points represent the result of five-fold cross-validation and the mean value of the cross-validation results, respectively. The diamond points indicate the test dataset results. Plots were created with the ggplot2 package in R version 4.1.1.^{28,29}

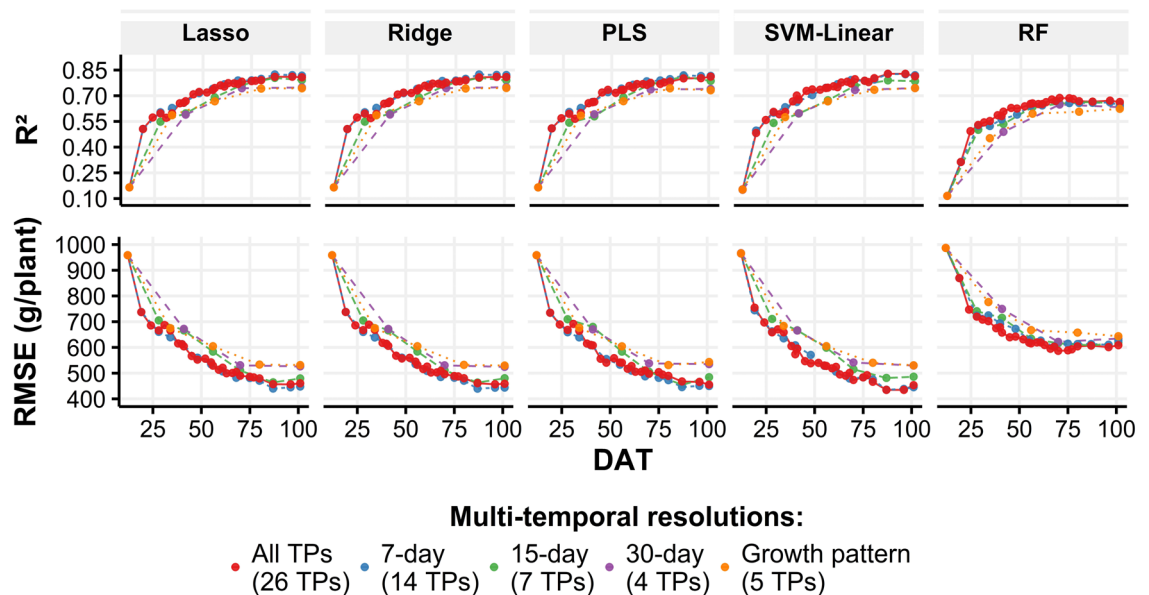


Figure 6. Early individual plant weight prediction. Prediction results (R^2 and RMSE) obtained from the test dataset ($n = 83$). The dots represent the results yielded by the regression models when these were trained using sequenced TPs data from 12 DAT up to the dot position. Each multi-temporal resolution is represented by different colors. Plots were created with the ggplot2 package in R version 4.1.1.^{28,29}

Discussion

In previous studies, UAV-based data extraction to characterize Chinese cabbage plants was based on either manually drawing a bounding box or using pixel-based classification methods^{20,21,30}. Although these methods are useful for obtaining crop phenotyping features at the canopy level, they hinder a more precise and efficient understanding at the individual plant level^{4,31}. In this study, an object detection algorithm automatically detects 95% of the Chinese cabbage plants sown in the field. In addition, the predicted bounding boxes facilitated the extraction of individual plant images captured by the two cameras (RGB and MS). Furthermore, individual plant images facilitated the computation of features, grouped into three levels (RGB-MS, VI, and 3D point-cloud data), throughout multiple TPs.

In this study, the weight prediction accuracy was evaluated under different combinations of UAV-derived features and multi-temporal resolutions. The most relevant features for predicting Chinese cabbage weight were identified from the three levels of UAV information. Regression models trained with a reduced number of UAV-derived features (seven) achieved accuracy predictions similar to those trained with all features (18) (Fig. 2), showing that using a large number of inputs for training is not efficient¹². Moreover, the results indicated that features from 3D point-cloud data, such as volume, achieved the highest prediction accuracy (Fig. 2), confirming the importance of including 3D-derived features for crop biomass prediction^{14,32}. Furthermore, combining UAV information from multiple sources (RGB, MS, VI, and 3D point cloud) resulted in an increase in weight prediction accuracy of 0.11 in R^2 and a decrease of 70 g/plant in RMSE compared to the one using only volume. These results indicated the effectiveness of using UAV information from multiple sources^{10,14}.

To optimize the number of TPs required for weight prediction, two issues were evaluated: the advantage of using multi-temporal data over a single time-point prediction and the prediction performance of regression models trained with lower multi-temporal resolution. For the first issue, implementing multi-temporal data led to an increase in the weight prediction accuracy in the test dataset in R^2 (0.17) and a decrease in RMSE (178 g/plant) compared to using only single time-point features (Fig. 4). This confirms previous findings on the use of multi-temporal remote sensing data for crop biomass prediction^{14,18}. Secondly, models using a lower multi-temporal resolution (14 TPs) yielded higher accuracies than those trained with all 26 TPs (Fig. 5). These results indicate that lower multi-temporal resolution (7-day and 15-day ones) can provide sufficient information for monitoring Chinese cabbage crops.

In addition to identifying the most relevant UAV features and multi-temporal resolutions for predicting Chinese cabbage weight, this study addressed the capability of predicting individual plant weights prior to harvest. By using sequenced TP data up to 53 DPH, the models started to achieve accuracies in R^2 greater than 0.72 and RMSE lower than 560 g/plant, reaching R^2 values of 0.8 and RMSE of 467 g/plant when data up to 21 DPH were used. This early prediction of weight can be explained by the fact that the weight of Chinese cabbage is linearly proportional to plant height²¹. The final plant size and shape were determined after the rosette stage³³ (Fig. S2). Our results are comparable to those reported by Kim et al.²¹ for a test dataset with R^2 values greater than 0.76. However, unlike their approach of predicting weight using RGB information at the time of harvest, our multi-temporal UAV-based data allowed us to predict weight with $R^2 > 0.8$ at 21 DPH. These results indicate that the use of multi-sourcing and multi-temporal UAV-based data significantly improves the accuracy of weight prediction and enables early weight prediction.

Conclusions

In this study, we explored the feasibility of predicting individual plant weights by automatically generating and optimizing multi-temporal UAV-based data. The proposed methodology integrates an object detection model that identifies plants within a field and computes the features from two cameras (RGB and MS) at the plant level. Furthermore, the effective use of relevant multi-temporal features to predict biomass enabled the prediction of Chinese cabbage weight at 53 DPH. The ability to predict early weight is important for farmers and breeders in crop management.

Materials and methods

The pipeline implemented in this study was divided into four main steps, as shown in Fig. 7.

Data acquisition

Plant materials and cultivation: Chinese cabbage cultivars were purchased, and their genetically fixed lines (P1 and P2) as well as their F2 populations were cultivated in a field at the Institute for Sustainable Agro-Ecosystem Services at the University of Tokyo, Nishi-Tokyo, Tokyo, Japan. Experimental research and field studies on plants, including the collection of plant materials, comply with relevant institutional, national, and international guidelines and legislation.

UAV imagery acquisition: The image-capturing process was conducted from September 22 to December 20, 2021, in an experimental field located in Tokyo, Japan. On each capture date, two UAVs were flown: a DJI Phantom 4 RTK and a DJI P4 Multispectral (SZ DJI Technology Co., Shenzhen, China). The DJI Phantom 4 RTK captured RGB images with a resolution of 5472×3648 pixels. The P4 Multispectral (MS) obtained imagery in five different bands: blue ($450 \text{ nm} \pm 16 \text{ nm}$), green ($560 \text{ nm} \pm 16 \text{ nm}$), red ($650 \text{ nm} \pm 16 \text{ nm}$), red edge ($730 \text{ nm} \pm 16 \text{ nm}$), and near-infrared (NIR) ($840 \text{ nm} \pm 26 \text{ nm}$), with a pixel resolution of 1600×1300 . Additionally, the flight altitude and image ratio overlap (front and side) were set to 15 m and 80%, respectively. To improve spatial accuracy, both UAVs were connected to a network RTK Service ("ichimill" SoftBank Corp., Tokyo, Japan). Four ground control points (GCPs) were placed at the field corners. Images were processed using the photogrammetric software Pix4D Mapper (version 4.6; Pix4D SA, Prilly, Switzerland)²³. This software was applied to create RGB and MS orthomosaic imagery and generate 3D point cloud data (Fig. 7). The configuration for Pix4D processing was set as described by Want et al.³⁴, as proposed in their study.

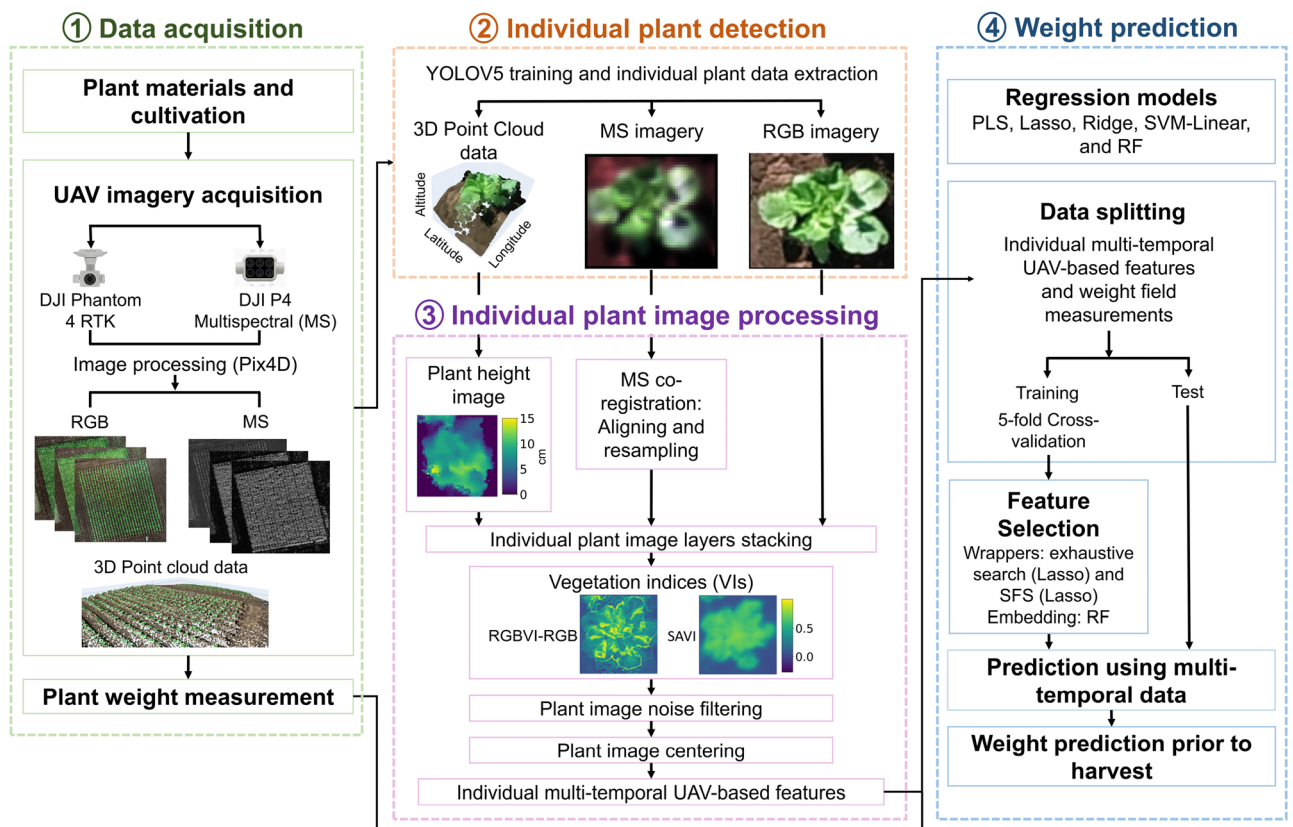


Figure 7. Workflow of the proposed methodology that uses the most relevant RGB, MS, and point cloud data to predict individual Chinese cabbage plant weight prior to harvest. The general framework is divided into four stages: data acquisition, individual plant detection, individual plant image processing, and weight prediction. Individual plant images obtained with Matplotlib package in Python version 3.9^{25,26}.

The individual plant weight was measured after 102 and 103 days after transplanting. The plants were cut off from the base and fresh weight was measured using a balance. To avoid errors in weighing, the plants were cut and placed in the order in which they were planted in the field and then weighed. This procedure was repeated for each row (46 plants/row) to ensure that the number of plants was accurate.

Individual plant detection

Individual Chinese cabbage plants were detected using YOLOv5 implemented in PyTorch³⁵. To facilitate labelling and model training, the RGB orthomosaic for each TP (September 29, October 4, and October 6) was divided into 512 × 512 tiles. For labelling, the tiles belonging to the training dataset were uploaded to the RoboFlow platform³⁶, where the bounding boxes were manually drawn. Using RGB orthomosaic imagery as a reference, we manually drew bounding boxes for 975 plants across the three TPs. To augment the training dataset, three transformation functions were randomly applied to each tile: rotation, hue-saturation-brightness modification, and expansion (Fig. S3). These functions were executed using the OpenCV-Python package³⁷.

To obtain a unique bounding box for each plant, the predicted bounding boxes obtained for each TP RGB orthomosaic image were merged. For this, bounding boxes whose intersected areas were greater than 40% were found and grouped. Subsequently, each group of intersecting bounding boxes was merged into a single bounding box by selecting the maximum height and width that covered all boxes. The merging step was performed using the GeoPandas library in Python³⁸.

To assign each plant data to its corresponding bounding box, we created a spatial point grid geometry using QGIS³⁹, with a spacing of 45 × 65 cm (distance between plants and rows). Each spatial point contained a unique index for plant identification. The spatial distance (SPd) from the point center (Pc) to every neighboring bounding box (BBc) was computed using Euclidean distance equation (Eq. 1). The closest bounding box to the spatial point with a distance less than 45 cm (plant distance sowing) was indexed to the point information.

$$SPd = \sqrt{(Pc_x - BBc_x)^2 + (Pc_y - BBc_y)^2} \quad (1)$$

The object detection performance was measured using the overall accuracy of the test field⁴⁰.

Individual plant image processing

To compute the features of each plant across the 26 TPs, the data derived from the RGB, MS, and 3D point clouds created by the RGB camera were processed, as shown in Fig. 7.

Plant-height image: To transform the 3D point cloud into a plant height image, the 3D point cloud was interpolated using neighboring values. To obtain an accurate plant height value, the reference surface value was subtracted from each plant-height image. The surface value was obtained from the first TP data (September 22, 2021) because the soil surface was not fully covered by plants. The spatial accuracy of the 3D point cloud data in the X- and Y- dimensions was assessed by comparing two RGB images: RGB imagery taken by RGB camera and 2D RGB image from 3D point cloud data. Two metrics were used: cross-correlation displacement⁴¹ and correlation-coefficient. Cross-correlation is a similarity metric, quantifying the displacement between the two images transformed into Fourier space. The assessment was conducted on a randomly selected 100 plants. For each plant, two types of RGB images (RGB camera and 2D RGB image from 3D point cloud data) were transformed into grayscale images. (Fig. S4). The maximum displacement in X and Y was 3.81 mm; the median of correlation coefficient was 0.89.

MS co-registration: To use both the RGB and MS imagery, the MS imagery was registered to the RGB resolution. Registration of the MS imagery was divided into two steps: alignment and resampling. The alignment step was performed using phase cross-correlation. This determines the linear shift of the two images by obtaining a correlation peak in their frequency domain representation, which is obtained by applying an inverse Fourier transform^{41,42}. MS resampling of the RGB spatial resolution was computed using a k-nearest neighbor interpolation.

Individual plant image layer stacking: To facilitate calculations across the multi-source image layers for all TPs, the three image sources (RGB, MS, and plant height images) were stacked into an individual plant image with four dimensions (TPs, layers, longitude, and latitude) using the Xarray package in Python^{26,43}.

Vegetation indices (VIs): Once the individual plant images were obtained, we calculated the VIs using the equations^{44–50} listed in Table 1. Initially, 13 VIs were calculated, but this number was reduced to seven after excluding highly correlated VIs (Pearson correlation coefficient > 0.95) (Fig. S5, red boxes).

Plant image noise filtering: Three filters were applied to each individual plant image to remove and reduce the noise caused by non-vegetation (soil) pixels, contours that were not neighboring the main plant, and pixel values affected by brightness and shadows (Fig. S6). The first two filters were applied to all the image layers in the individual plant image, whereas the third filter was applied only to the RGB layers. For the first filter, a soil mask was obtained from a k-means model trained using seven VIs. The k-means model was implemented using the scikit-learn package. To implement the second filter, the OpenCV-Python package³⁷ was used to compute contours wrapping neighboring pixels, and a filter based on their area was then applied to remove smaller contours. For the third filter, a histogram equalization technique⁵¹ was applied to each RGB layer. A histogram was constructed using the Numpy package⁵².

Plant image centering: To center the individual plant images, a convex hull polygon was computed using the first TP RGB plant image (September 22, 2021). The geometric center of the polygon was obtained by averaging its vertices, and each plant image layer was then shifted to the center of the polygon. The SciPy package was used to obtain convex hulls⁵³. Finally, the image was reduced by 70% from the center to the border (Fig. S6).

| Vegetation Indices (VIs) | Formula | References | Camera |
|---|---|------------|------------|
| GRVI: Green and red ratio vegetation index | $\frac{\text{green}-\text{red}}{\text{green}+\text{red}}$ | 44 | RGB and MS |
| MGRVI: Modified green-red vegetation index | $\frac{\text{green}^2-\text{red}^2}{\text{green}^2+\text{red}^2}$ | 45 | RGB and MS |
| RGBVI: Red-green-blue vegetation index | $\frac{\text{green}^2-(\text{blue}*\text{red})}{\text{green}^2+(\text{blue}*\text{red})}$ | 45 | RGB and MS |
| GNDVI: Green normalized difference vegetation index | $\frac{\text{NIR}-\text{green}}{\text{NIR}+\text{green}}$ | 46 | MS |
| NDRE: Normalized difference red-edge index | $\frac{\text{NIR}-\text{red_edge}}{\text{NIR}+\text{red_edge}}$ | 47 | MS |
| NDVI: Normalized difference vegetation index | $\frac{\text{NIR}-\text{red}}{\text{NIR}+\text{red}}$ | 48 | MS |
| RECI: Red-edge chlorophyll index | $\frac{\text{NIR}}{\text{red_edge}} - 1$ | 49 | MS |
| SAVI: Soil-adjusted vegetation index | $\frac{\text{NIR}-\text{red}}{\text{NIR}+\text{red}+0.5} * 1.5$ | 50 | MS |

Table 1. The computed vegetation indices. Blue, green, red, near-infrared (NIR), and red-edge represent the image layer values.

Individual multi-temporal UAV-based features: Individual plant image layers were summarized as features. The RGB and MS layers were summarized using the median statistic (Fig. S7) and the plant height image was transformed into three features: plant height (PH), leaf angle, and volume. The PH of each TP was computed as the 90th percentile of plant height image values (Fig. S8). The leaf angle was calculated as the angle between each pixel vector (P_p) (represented as longitude, latitude, and altitude) and a normal vector located in the plant height image center (V_c); the distance between the two vectors was used to compute the angle (Eq. 2). The individual plant volume was calculated as the sum of all pixel height values (H_i), which comprised the 2D plant height image, and was then multiplied by the pixel area in cm^2 (a_p) (Eq. 3)⁴.

$$\cos(\text{leaf angle}) = \frac{P_p V_c}{\|P_p\| \|V_c\|} \quad (2)$$

$$\text{Volume} = \sum_{i=1}^n H_i * a_p \quad (3)$$

Finally, to train the regression models, the features were standardized using the StandardScaler function from the scikit-learn library⁵⁴.

Weight prediction

Regression models: Model hyperparameters were set through an exhaustive search using a five-fold cross-validation approach. The hyperparameters used to determine the optimal configuration for each model are listed in Table S3. Scikit-learn was used to implement and optimize the models.

Data splitting: To evaluate the weight prediction results, individual plant data were split into two groups: training (90%) and testing (10%) datasets. The training dataset was split into 5-folds to implement cross-validation (Fig. 7).

Feature selection: The lasso regression model was used to find the best input combination for both wrapper methods. For an exhaustive search, the total number of combinations was configured according to the total number of features available at each level of information. SFS was implemented using the Sequential Feature Selector function available in the scikit-learn library, and forward selection was set as the search method. The embedding method was applied through the RF, in which the mean decrease in impurity (MDI) was used to rank the feature importance.

Evaluation metrics

To assess the accuracy of individual weight predictions, two evaluation metrics were implemented: the coefficient of determination (R^2) (Eq. 4) and the root mean square error (RMSE) (Eq. 5).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5)$$

where, y_i is the individual plant weight, \bar{y} is the average weight, and \hat{y}_i is the individual plant weight predicted using the model.

Data availability

The UAV imagery generated and/or analyzed during the current study is available from the corresponding author upon reasonable request.

Code availability

All code associated with the current study is available at: https://github.com/anaguilarar/CC_Weight_Prediction.

Received: 13 June 2023; Accepted: 14 November 2023

Published online: 17 November 2023

References

- Bisbis, M. B., Gruda, N. & Blanke, M. Potential impacts of climate change on vegetable production and product quality—a review. *J. Clean. Prod.* **170**, 1602–1620 (2018).
- Ray, D. K., Gerber, J. S., MacDonald, G. K. & West, P. C. Climate variation explains a third of global crop yield variability. *Nat. Commun.* **6**, 5989 (2015).
- Song, P., Wang, J., Guo, X., Yang, W. & Zhao, C. High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *Crop J.* **9**, 633–645 (2021).
- Guo, W., Fukano, Y., Noshita, K. & Ninomiya, S. Field-based individual plant phenotyping of herbaceous species by unmanned aerial vehicle. *Ecol. Evol.* **10**, 12318–12326 (2020).
- Deng, L. *et al.* UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. *ISPRS J. Photogram. Remote Sens.* **146**, 124–136 (2018).
- Yang, G. *et al.* Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* **8**, 1111. <https://doi.org/10.3389/fpls.2017.01111> (2017).
- Guo, W. *et al.* UAS-based plant phenotyping for research and breeding applications. *Plant Phenom.* **2021**, 9840192. <https://doi.org/10.34133/2021/9840192> (2021).
- Bannari, A., Morin, D., Bonn, F. & Huete, A. R. A review of vegetation indices. *Remote Sens. Rev.* **13**, 95–120 (1995).
- Tang, Z. *et al.* Validation of UAV-based alfalfa biomass predictability using photogrammetry with fully automatic plot segmentation. *Sci. Rep.* **11**, 1. <https://doi.org/10.1038/S41598-021-82797-X> (2021).
- Maimaitijiang, M. *et al.* Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* **237**, 111599. <https://doi.org/10.1016/j.rse.2019.111599> (2020).
- Ghamisi, P. *et al.* Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **7**, 1. <https://doi.org/10.1109/MGRS.2018.2890023> (2019).
- Barbosa, B. D. S. *et al.* UAV-based coffee yield prediction utilizing feature selection and deep learning. *Smart Agric. Technol.* **1**, 100010. <https://doi.org/10.1016/j.atech.2021.100010> (2021).
- Li, B. *et al.* Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS J. Photogram. Remote Sens.* **162**, 161–172 (2020).
- Ji, Y. *et al.* Estimation of plant height and yield based on UAV imagery in faba bean (*Vicia faba* L.). *Plant Methods* **18**, 26. <https://doi.org/10.1186/s13007-022-00861-7> (2022).
- Fei, S. *et al.* UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. *Precis. Agric.* <https://doi.org/10.1007/s11119-022-09938-8> (2022).
- Feng, A., Zhou, J., Vories, E. D., Sudduth, K. A. & Zhang, M. Yield estimation in cotton using UAV-based multi-sensor imagery. *Biosyst. Eng.* **193**, 101–114 (2020).
- Ashapure, A. *et al.* Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data. *ISPRS J. Photogram. Remote Sens.* **169**, 180. <https://doi.org/10.1016/j.isprsjprs.2020.09.015> (2020).
- Nevavuori, P., Narra, N., Linna, P. & Lipping, T. Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models. *Remote Sens. Basel* **12**, 23. <https://doi.org/10.3390/rs12234000> (2020).
- Sun, X. X. *et al.* Genetic analysis of Chinese cabbage reveals correlation between rosette leaf and leafy head variation. *Front. Plant Sci.* **9**, 1455. <https://doi.org/10.3389/fpls.2018.01455> (2018).
- Kang, Y.-S. *et al.* Yield prediction of Chinese cabbage (*Brassica rapa* var. *glabra* Regel.) using narrowband hyperspectral imagery and effective accumulated temperature. *J. Agric. Life Sci.* **54**, 3. <https://doi.org/10.14397/jals.2020.54.3.95> (2020).
- Kim, D. W. *et al.* Modeling and testing of growth status for Chinese cabbage and white radish with UAV-based RGB imagery. *Remote Sens. Basel* **10**, 4. <https://doi.org/10.3390/rs10040563> (2018).
- Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2016).
- PIX4D. Professional photogrammetry and drone mapping software | Pix4D. *PIX4D*. <https://www.pix4d.com/> (2021).
- Esri Inc. ArcGIS Pro (Version 3.0.3). *Esri Inc.* <https://www.esri.com/en-us/home> (2023).
- Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 3. <https://doi.org/10.1109/MCSE.2007.55> (2007).
- Phillips, D. *Python3 Object-oriented Programming*, vol. 58 12. <https://doi.org/10.1109/TGRS.2004.834800> (2014).
- Chandrashekar, G. & Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **40**, 16–28 (2014).
- Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
- R Core Team. *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/> (2021).
- Zhang, J. *et al.* Multispectral drone imagery and SRGAN for rapid phenotypic mapping of individual chinese cabbage plants. *Plant Phenom.* **2022**, 896. <https://doi.org/10.34133/plantphenomics.0007> (2022).
- Pantazi, E. *et al.* Assessment of different object detectors for the maturity level classification of broccoli crops using UAV imagery. *Remote Sens. Basel* **14**, 3. <https://doi.org/10.3390/rs14030731> (2022).
- Fu, H., Wang, C., Cui, G., She, W. & Zhao, L. Ramie yield estimation based on UAV RGB images. *Sens. Switzerl.* **21**, 2. <https://doi.org/10.3390/s21020669> (2021).
- Sun, X. X. *et al.* Genome-wide transcriptome analysis reveals molecular pathways involved in leafy head formation of Chinese cabbage (*Brassica rapa*). *Hortic. Res.* **6**, 130. <https://doi.org/10.1038/s41438-019-0212-9> (2019).
- Wang, H. *et al.* EasyIDP: A python package for intermediate data processing in UAV-based plant phenotyping. *Remote Sens. Basel* **13**, 13. <https://doi.org/10.3390/rs13132622> (2021).
- Paszke, A. *et al.* *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. <https://pytorch.org/> (2019).
- Dwyer, B. & Nelson, J. Roboflow (Version 1.0) [Software]. <https://roboflow.com/> (2022).
- Bradski, G. The OpenCV Library. *Dr. Dobbs' Journal of Software Tools*. <https://opencv.org/> (2000).
- Jordahl, K. GeoPandas: Python tools for geographic data. <https://github.com/geopandas/geopandas> (2014).
- QGIS Development Team. QGIS Geographic Information System. <http://qgis.org> (2009).
- Sokolova, M. & Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **45**, 427–437 (2009).
- Kuglin, C. D. & Hines, D. C. The phase correlation image alignment method. *IEEE Int. Conf. Cybern. Soc.* **6**, 163–165 (1975).
- Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett.* **33**, 156–158 (2008).
- Hoyer, S. & Hamman, J. xarray: N-D labeled arrays and datasets in Python. *J. Open Res. Softw.* **5**, 1. <https://doi.org/10.5334/jors.148> (2017).

44. Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **8**, 127–150 (1979).
45. Bendig, J. *et al.* Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Observ. Geoinf.* **39**, 79–87 (2015).
46. Gitelson, A. A., Gritz, Y. & Merzlyak, M. N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **160**, 271–282 (2003).
47. Maccioni, A., Agati, G. & Mazzinghi, P. New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. *J. Photochem. Photobiol. B* **61**, 52–61 (2001).
48. Rouse, J. W., Haas, R. H., Schell, J. A. & Deering, D. W. Monitoring vegetation systems in the great plains with ERTS. In *Third Earth Resources Technology Satellite-1 Symposium, vol. 1* (1973).
49. Gitelson, A. & Merzlyak, M. N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol. B* **22**, 247–252 (1994).
50. Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **25**, 295–309 (1988).
51. Hum, Y. C., Lai, K. W. & Mohamad Salim, M. I. Multiobjectives bihistogram equalization for image contrast enhancement. *Complexity* **20**, 2. <https://doi.org/10.1002/cplx.21499> (2014).
52. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
53. Virtanen, P. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
54. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Acknowledgements

This work was supported by the Cabinet Office, the Government of Japan, and the Moonshot Research and Development Program for Agriculture, Forestry, and Fisheries (funding agency: Bio-oriented Technology Research Advancement Institution).

Author contributions

A.A., T.K., W.G., and T.F. designed the experiments. A.A. conducted the data analysis with significant contributions from T.K., W.G., and T.F. M.I. and W.G. captured the UAV data. A.A., T.K., T.M., A.S., H.P.K., H.W.P., and T.M. collected the field data. A.A. and T.K. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-47431-y>.

Correspondence and requests for materials should be addressed to T.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023