Smart Agricultural Technology 11 (2025) 100998

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Robustness of the RGB image-based estimation for rice above-ground biomass by utilizing the dataset collected across multiple locations

Kota Nakajima [a,b], Kazuki Saito [c], Yasuhiro Tsujimoto [d], Toshiyuki Takai [d], Atsushi Mochizuki [e], Tomoaki Yamaguchi [f], Ali Ibrahim [g], Salifou Goube Mairoua [h], Bruce Haja Andrianary [i], Keisuke Katsura [a], Yu Tanaka [j,*]

[a] Graduate School of Agriculture, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
[b] Japan Society for the Promotion of Science, Tokyo, Japan
[c] International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila 1301, Philippines
[d] Japan International Research Center for Agricultural Sciences, 1-1 Ohwashi, Tsukuba, Ibaraki, 3058686, Japan
[e] CHIBA Prefectural Agriculture and Forestry Research Center, 180-1 Okanezawa-cho, Midori-ku, Chiba, Chiba, 266-0014, Japan
[f] Faculty of Applied Biological Sciences, Gifu University, Yanagido, Gifu, 5011193, Japan
[g] Africa Rice Center (AfricaRice), Regional Station for the Sahel, B.P. 96, Saint-Louis, Senegal
[h] Africa Rice Center (AfricaRice), 01 BP 2551, Bouaké, Côte d'Ivoire
[i] Laboratoire des Radioisotopes, Université d'Antananarivo, Route d'Andraisoro, 101, BP 3383, Madagascar
[j] Graduate School of Environment, Life, Natural Science and Technology, Okayama University, 3-1-1 Tsushimanaka, Kita-ku, Okayama 700-8530, Japan

ARTICLE INFO

Keywords:
Robustness
RGB image
Rice, Above-ground biomass
Convolutional neural network

ABSTRACT

Above-ground biomass (AGB) is a critical phenotype representing crop growth. Non-invasive evaluations of AGB, including deep-learning-based red-green-blue (RGB) image analyses, are often specific to the training data. The robustness of the estimation model across untrained conditions is essential to monitor crop productivity globally, but it has yet to be fully assessed. This study aims to assess the robustness of a convolutional neural network (CNN) model for rice AGB estimation across five locations in three countries, and to demonstrate the feasibility of robust model via a practical approach. From transplanting to heading, 1957 RGB images were captured vertically downward over the rice canopy, covering approximately 1 m^2. First, a base model was established using data collected from a single location. Then, its robustness was assessed using test datasets taken from the other four locations. The CNN model showed a significant variation in estimation accuracy across the untrained four locations, indicating insufficient robustness of the base model. Subsequently, we quantitatively tested the impact of improving training data diversity on model robustness by adding data from each of the four locations to the base model's training data. Adding at most 48 data points from a location achieved practical accuracy for the added location, with R^2_{Ad} above 0.8. Interestingly, adding data from one location sometimes improved the accuracy for other untrained locations as well. These findings suggest that collecting diverse training data for RGB-based estimation, combined with evaluation of robustness paves the way for on-site and instant AGB monitoring of rice.

1. Introduction

Crop production is a fundamental part of human society for ensuring food security. As the global population continues to grow, the demand for crop production increases accordingly [1]. Above-ground biomass (AGB), defined as the weight of above-ground dried plants per unit area, is one of the critical phenotypes for assessing crop growth. AGB reflects the cumulative outcome of complex physiological processes in crops, including photosynthesis and nutrient absorption [2], thereby representing crop growth up to a given point. Moreover, AGB after the heading stage is closely related to crop yield [3]. Thus, AGB plays a key role in monitoring crop growth and predicting crop productivity [4]. Rice (*Oryza Sativa* L.), one of the world's major food crops [5], exhibits significant genetic diversity, and its growth conditions are also

* Corresponding author.
E-mail addresses: nakajima.kota.65m@st.kyoto-u.ac.jp (K. Nakajima), k.saito@irri.org (K. Saito), tsujimotoy0365@jircas.go.jp (Y. Tsujimoto), toshi336@affrc.go.jp (T. Takai), a.mchdk5@pref.chiba.lg.jp (A. Mochizuki), yamaguchi.tomoaki.t0@f.gifu-u.ac.jp (T. Yamaguchi), I.Ali@cgiar.org (A. Ibrahim), hajabruce@yahoo.fr (B.H. Andrianary), katsura.keisuke.2a@kyoto-u.ac.jp (K. Katsura), yutanaka@okayama-u.ac.jp (Y. Tanaka).

https://doi.org/10.1016/j.atech.2025.100998
Received 20 February 2025; Received in revised form 7 May 2025; Accepted 10 May 2025
Available online 13 May 2025

incredibly diverse [6]. It is well known that AGB is greatly influenced by both genetic and environmental factors [7,8]. Therefore, measuring AGB across a wide range of genetic variations and growth conditions is crucial for improving crop productivity.

Conventionally, AGB has been measured destructively, requiring manual harvesting, drying, and weighing. This process is time-consuming and labor-intensive, making extensive AGB measurements across varied conditions impractical. Limited access to AGB data has been cited as a bottleneck for both genetic improvements for AGB productivity and the assessment of AGB productivity at field scale under varied conditions [9,10]. Consequently, there is a need for a high-throughput, robust and user-friendly method to monitor rice AGB across diverse varieties and growth conditions.

Numerous studies on AGB estimation have significantly contributed to increasing estimation accuracy and expanding the applicable range of growth stages. Various sensor types, input data, and regression methods have been widely employed for AGB estimation [11,12]. For instance, sensors such as RGB, multispectral, and hyperspectral imaging, as well as Light Detection and Ranging and Synthetic Aperture Radar, have been examined [13–15]. In terms of input data, both raw sensing data, such as RGB images, and derived data calculated from sensing data, such as vegetation indices, estimated plant height, texture, and 3D point cloud, have been shown to be effective for AGB estimation [14,16–19]. Additionally, other data sources, such as meteorological data and phenology information, have also been found to contribute to AGB estimation [18, 20]. Regarding regression methods, approaches including linear and multiple regression, machine learning techniques such as random forests and artificial neural network, as well as deep learning methods like convolutional neural network (CNN), have been extensively explored [13,17,19,21,22]. However, a robust AGB estimation method that can be widely applied across the world remains elusive—in other words, establishing a method that is both easy-to-use and robust to diverse varieties and growth conditions remains a critical goal in this field.

Robustness of AGB estimation methods, especially their performance in conditions not represented in training data, have not been fully evaluated in previous studies. While our previous study showed that a CNN model trained on genetically diverse rice varieties could accurately estimate AGB even for varieties not included in the training process [25], robustness to new locations remains uncertain Previous studies typically assessed robustness of the model to multiple locations through cross-validation using spatially interpolated data from the same locations as training data [19,23,24]. While these methods avoid data leakage, they fail to evaluate model performance for spatially extrapolated data, which is critical for assessing the model's robustness to new locations.

To address this gap, we evaluated the robustness of a CNN-based AGB estimation using data collected from five geographically distinct locations in three countries. Because collecting ground-truth biomass is costly and our dataset size is therefore limited, we adopted a CNN, whose local-feature bias allows reliable training with limited data [26]. First, we established a base model trained on data from a single location. To accurately evaluate the robustness of the CNN model across different locations, it was necessary to minimize biases due to varietal differences. Therefore, we intentionally selected morphologically diverse rice varieties, including near-isogenic lines differing primarily in plant architecture, to establish a base model having robustness to varieties. Then, we evaluated its basic performance with test data from the trained location and then its robustness using data from four independent locations not included in the training dataset. Subsequently, we incrementally added data points from each location to the training dataset. We aimed to clarify not only whether accuracy improved at the added locations, but also whether adding data from one location could enhance estimation accuracy at other unseen locations. Crucially, our study quantitatively determines, for the first time, the minimal amount of data required from a new location to achieve practical accuracy, providing valuable insights for efficient data collection strategies. Our findings reveal a promising pathway toward developing truly robust and universally applicable AGB estimation methods.

2. Materials and methods

2.1. Cultivation conditions

Rice was cultivated across five locations, three locations in Japan (Kyoto, Tokyo, and Chiba), one location in Côte d'Ivoire and one location in Madagascar (Supplemental Fig. S1). In Kyoto, rice was cultivated in 2021 at the experimental paddy field in the Graduate School of Agriculture, Kyoto University, to obtain a training dataset comprising morphologically diverse rice varieties for the base model. Seeds were sown on April 27th, and seedlings were transplanted on May 20th. Eight varieties were cultivated at a density of 22.2 hills m^{-2} (15 cm between hills, 30 cm between rows). Among eight varieties, a temperate *japonica* variety, Koshihikari was selected as the representative variety in Japan. An *indica* variety, Takanari, was cultivated as a high-yielding variety. Two near-isogenic lines with a genetic background of the Takanari; Takanari-*SD1* and Takanari-*tac1* were cultivated. Takanari-*SD1* carries a functional allele of *Semi Dwarf1* (*SD1*, [27]) derived from Koshihikari, which increased height of Takanari. Takanari-*tac1* carries a low-function allele of *Tiller Angle Control 1* (*TAC1*, [28]) derived from Koshihikari, which erected tillers of Takanari. The remaining four varieties, Ma sho, Jinguoyin, Co 13, and Khao Nok, were selected from the World Rice Core Collection [29] to cover the morphological diversity based on our previous research [25]. The intentional selection of these diverse varieties, including NILs, aimed to minimize bias due to varietal differences affecting estimation accuracy when evaluating the model's robustness across different locations.In Tokyo, cultivation was conducted in 2021 at the Field Museum Honmachi, Tokyo University of Agriculture and Technology. Koshihikari was sown twice at approximately 40-day intervals and transplanted after about two weeks of nursery growth. In Chiba, cultivation was conducted in 2022 and 2023 at the Chiba Prefectural Agriculture and Forestry Research Center. Koshihikari and two other temperate *japonica* varieties, Fusaotome and Fusakogane, were sown in five batches between March 11th and May 20th in 2022, and between March 14th and May 19th in 2023, and transplanted after two weeks to one month of nursery growth. In Côte d'Ivoire, cultivation took place in 2020, 2021, and 2022 at the AfricaRice M'bé Research Station. Up to four varieties were cultivated each year, either by transplanting or direct seeding. In Madagascar, cultivation was conducted in 2020 and 2021 on farmers' fields in Behenjy, Betafo, and Faratsiho which are all located in the central highlands. In 2020, three local varieties, and in 2021, one variety were transplanted. Details on sowing and transplanting dates, varieties, water management, and total fertilizer at each location are provided in Supplemental Table S1.

2.2 Dataset construction

The plot for taking RGB images and measuring corresponding AGB was set as approximately 1 m^2 (90 cm x 120 cm). The canopy image and its corresponding AGB were collected from two weeks after transplanting up to the heading stage. Images with a 3:4 aspect ratio were taken vertically downward from above the canopy. Camera types varied by location and year (Supplemental Table S2). In Kyoto, images were captured from approximately 3 m above the canopy using a small drone (Mavic Mini, DJI, China), allowing efficient capture of a lot of plots. Subsequently, the central one-third portion of the image was cropped (Supplemental Fig. S2). At other locations, images were captured from approximately 90 cm above the canopy using commercially available digital cameras or smartphones (Supplemental Fig. S3). After taking images, all above-ground parts of the rice plants in the plot were harvested. The harvested plants were dried in an oven at 80 °C for over 48 h, and their weight was measured to calculate AGB (t ha^{-1}).

In Kyoto, the assignment of each plot to a specific dataset was

predetermined. To efficiently construct the training dataset, five overlapping plots were created by shifting each plot by the width of one plant (Supplemental Fig. S4). In contrast, the validation and test datasets were collected from non-overlapping plots to ensure independent data collection. In total, 497 data points were collected, with the training, validation, and test datasets allocated in a 5:1:1 ratio. These datasets were consistently used in both the base model and additional data models.

In other locations, data was collected without overlapping. The data were managed separately by location, and all data from each location was divided into two parts with different intended uses. The first part, comprising 50 data points, was consistently used as the test dataset for both the base model and additional data models. The second part contains the remaining data and was used as additional data for training and validation datasets to improve the robustness of the model. To ensure that the test dataset from each location is representative of its corresponding location, we used the Kullback-Leibler divergence (KLD) to measure the similarity between two probability distributions, denoted by P and Q, as follows:

$$KLD(P||Q) = \sum_i P_i log \frac{P_i}{Q_i} \quad (1)$$

In this equation, P_i represents the probability distribution of AGB within each part (either the test dataset or the subset of remaining data) for a given location and Q_i denotes the AGB probability distribution over the entire dataset for that location. The data at each location was randomly divided 1000 times. For each division, we calculated two KLD values: (1) between the test dataset and the entire data from that location, and (2) between the remaining part and the entire data from that location. We then selected the division with the smallest difference between these two KLD values, thus optimizing the similarity of the test dataset to the entire dataset at each location.

In total, 1957 data points were collected across the five locations. The distribution of the observed AGB in each dataset and overview of each dataset are shown in Supplemental Fig. S5 and Table 1, respectively. All images were cropped to 1:1 aspect ratio by equally trimming the long sides of the image and then resized to 512 × 512 pixels. Consequently, the ground sample distance (GSD) for each input image was approximately 0.18 cm per pixel. Three types of offline augmentation—vertical flipping, horizontal flipping, and brightness adjustment by ±0.2—were equally applied to each image in the training dataset that increased the total number of images by twelve-fold (Supplemental Fig. S6).

2.3 Training of the base model

The model network was adopted from our previous research on rice yield estimation (Supplemental Fig. S7, [30]). This network is characterized by a significantly smaller number of parameters compared to existing networks, such as ResNet [31] and DenseNet [32]. The loss function was the root mean square error (RMSE) (Eq. (2)), defined as

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

Table 1
Overview of the dataset used for this research.

Dataset	Location	No. of images
Training	Kyoto	355
Validation	Kyoto	71
Test	Kyoto	71
Test and additional data, managed by location	Tokyo	110
	Chiba	410
	Côte d'Ivoire	710
	Madagascar	230

where n is the number of data, y_i and $\hat{y}_i$ are the estimated and observed AGB, respectively. The optimizer was Adam [33], and the number of epochs was set to 500. The validation loss was calculated at each epoch. The model weights were saved at the lowest validation loss during training, with early stopping set at 100 epochs. The learning rate was initialized and set to decay progressively throughout training, following an exponential decay schedule. A grid search was performed to find the optimal initial learning rate and batch size. Learning rates (0.1, 0.01, 0.001, 0.0001) and batch sizes (16, 32, 64, 128) were tested, each with five replications. For each combination, the minimum validation losses from the five replications were averaged, and the hyperparameters with the lowest average were selected as optimal. Among the five replicate models trained under the optimal hyperparameters, the model with the lowest validation loss was selected as base model for further evaluation. Training and evaluation were conducted using two RTX 3090 GPUs on a Linux OS (Ubuntu 18.04). The model was implemented using Python 3.10.0 (http://www.python.org) and TensorFlow 2.7.0 [34].

2.4. Evaluation of the base model

The performance of the base model was evaluated by inputting images from each dataset. The adjusted coefficient of determination (R^2_{Ad}) (Eq. (3)) and relative RMSE (rRMSE) (Eq. (4)) were used as indicators of the estimation accuracy.

$$R^2_{Ad} = 1 - \left(\frac{(1-R^2) \times (n-1)}{n-p-1}\right) \quad (3)$$

where R^2 is the coefficient of determination before the adjustment. n means the number of data, and p indicates the number of explanatory variables, respectively.

$$rRMSE = \frac{1}{\bar{y}} RMSE \quad (4)$$

where $\bar{y}$ is the average of the observed AGB.

The model's basic performance was evaluated using the validation and test datasets from Kyoto. Subsequently, the robustness of the model was assessed with the test datasets from four independent locations.

2.5. Incremental addition of data collected from a new location to the training data

We added data from the four locations other than Kyoto to the base model's training and validation datasets. We then trained the model and evaluated its robustness (Supplemental Fig. S8). To assess the robustness of the newly trained models on untrained locations even after data addition, additional data were selected from only one of the four locations other than Kyoto. By varying the location from which the additional data were drawn, four distinct data addition scenarios were implemented. The amount of additional data was incrementally increased to determine how many data points are needed to achieve sufficient accuracy for the new location. For each location, all data excluding the test dataset were incrementally added to the training and validation dataset in a 5:1 ration. Specifically, data points were added to the training dataset in increments of 10 from 10 to 50, and then in increments of 50 thereafter. The KLD-based method was also used to decide the pattern of division of additional data. To minimize the impact of data bias on the result, five different patterns were created to serve as replications. The model was trained with each dataset. The same data augmentation was applied as described in Section 2.2, "Dataset Construction." The optimal initial learning rate and batch size, identified in Section 2.3, "Model training," were used, while all other training settings were kept the same. Finally, after training with each dataset, the models were evaluated using the test dataset from Kyoto and other four locations.

3. Results

The optimal batch size and initial learning rate were explored using a grid search. Fig. 1(a) shows the average validation loss from five replications for each combination. The lowest validation loss, 0.378, was observed when the batch size was 32 and the initial learning rate was 0.001. The highest validation loss, 0.897, occurred with a batch size of 16 and an initial learning rate of 0.1. Under the optimal combination of hyperparameters, the model with the lowest validation loss among five replications was selected as base model and used for the subsequent evaluations. Fig. 1(b) shows the learning curve of the base model. The lowest validation loss of 0.342 was recorded at the 250th epoch. The early stopping resulted in the training process terminating at the 350th epoch.

Fig. 2(a, b) shows the estimation results of the base model for the validation and test dataset, respectively. For the validation dataset, R^2_{Ad} was 0.98, and rRMSE was 0.10. For the test dataset, R^2_{Ad} was 0.96 and rRMSE was 0.17. Supplemental Table S3 and S4 provide variety-wise estimation accuracy for the validation and test datasets, respectively. In the validation and test datasets, lowest R^2_{Ad} were 0.97 for Ma sho and 0.93 for Jinguoyin, respectively.

Fig. 3 displays the estimation accuracy of the base model on the test dataset for four independent locations. The R^2_{Ad} for Tokyo, Chiba, Côte d'Ivoire, and Madagascar were 0.65, 0.57, 0.75, and −0.37, respectively. The rRMSE for the four locations were 0.48, 0.57, 0.45, and 1.27, respectively. There was a significant difference in estimation accuracy between locations. In Chiba and Madagascar, there was a clear overestimation when observed AGB was approximately <4 t ha^{-1}. In Tokyo and Chiba, underestimation occurred when observed AGB was approximately >6 t ha^{-1}.

Fig. 4, Supplemental Fig. S9, and Supplemental Table S5 show the estimation results for five locations, with data added individually for each location. The estimation accuracy for Kyoto (test data) remained nearly unchanged compared to before data addition ($R^2_{Ad} = 0.96$, rRMSE $= 0.17$). In contrast, accuracy improved significantly for locations where data were added as the amount of added data increased. Adding at most 40 data points from a given location to the training dataset achieved an R^2_{Ad} above 0.8 for that location, including Madagascar, where estimation accuracy was notably lower before data addition (Fig. 4, Supplemental Fig. S9). Fig. 5 displays the comparison of estimation accuracy for the five locations before and after adding 40 data points from each location. Fig. 5 shows that adding 40 data points significantly corrected the underestimation observed in Tokyo and Chiba, and the overestimation in Chiba and Madagascar, leading to improved estimation accuracy. Interestingly, accuracy for some locations also improved even when data from other locations were added (Figs. 4, 5). In particular, estimation accuracy for Tokyo and Côte d'Ivoire generally improved with the addition of data from other locations, finally achieving an R^2_{Ad} above 0.8. In contrast, no significant improvement was observed for Chiba and Madagascar when data from other locations were added.

4. Discussion

4.1. Establishment of the base model trained on data only from Kyoto

We established the base model trained on data from Kyoto. Through the grid search-based optimization, we observed significant differences in validation loss across various hyperparameter settings, with a twofold difference between the lowest and highest losses (Fig. 1(a)). Furthermore, the base model's accuracy on validation and test datasets was comparable to or exceeded that of existing CNN-based AGB estimation models (Fig. 2, [14,25]). These results indicate that the model was a well-established and reliable baseline for assessing robustness across multiple locations.

4.2. Robustness of the base model across multiple locations

The base model showed significant differences in estimation accuracy across the four independent locations (Fig. 3). Even in Côte d'Ivoire, where the highest R^2_{Ad} (0.75) was achieved, the rRMSE was high at 0.45. This suggests that a model trained on data from a single location lacks sufficient robustness for other locations. Cultivation conditions in the four locations other than Kyoto differed from those in Kyoto in terms of planting density, crop establishment methods, and levels of applied fertilizer (Supplemental Table S1). Although the coverage area and GSD of the input image were standardized across locations, the person and cameras for taking images varied by location. Additionally, images from Kyoto were taken at a higher altitude using a drone, followed by cropping, unlike the other locations (Supplemental Fig. S2, Supplemental Table S2). These factors might cause differences in visual features in the images at each location, likely contributing to the large estimation errors in the base model at these locations.

The observed limitations of the base model align with findings from previous research. Habibi et al. [35] reported that estimation accuracy for spatial extrapolation data tends to be lower than for interpolation data in remote sensing-based soybean yield prediction. Similarly, Ferraciolli et al. [36]. highlighted the presence of spatial autocorrelation between observations, which can impact the reliability of predictions when applied to unseen locations. These studies underline the inherent challenges of achieving high accuracy in models trained on

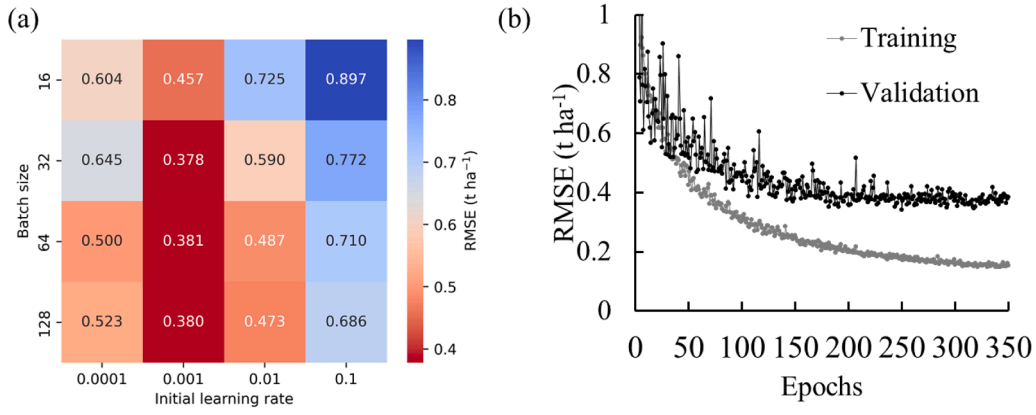

Fig. 1. Hyperparameter tuning results and learning curve. (a) Heatmap of validation loss (measured as RMSE) for different combinations of batch size and initial learning rate, averaged over five training replications. The x-axis represents the initial learning rate, and the y-axis represents the batch size. (b) Learning curve of the model achieving the lowest validation loss among five individually trained models. This model was trained on Kyoto data with the optimal combination of batch size and initial learning rate, as determined in (a), showing both training and validation loss across epochs.

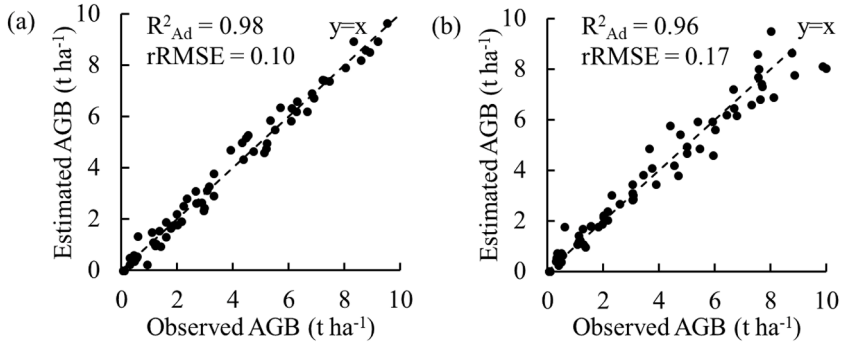

Fig. 2. Estimation results of the base model on the (a) validation and (b) test dataset. The dotted line represents the 1:1 line, indicating perfect agreement between observed and predicted values.

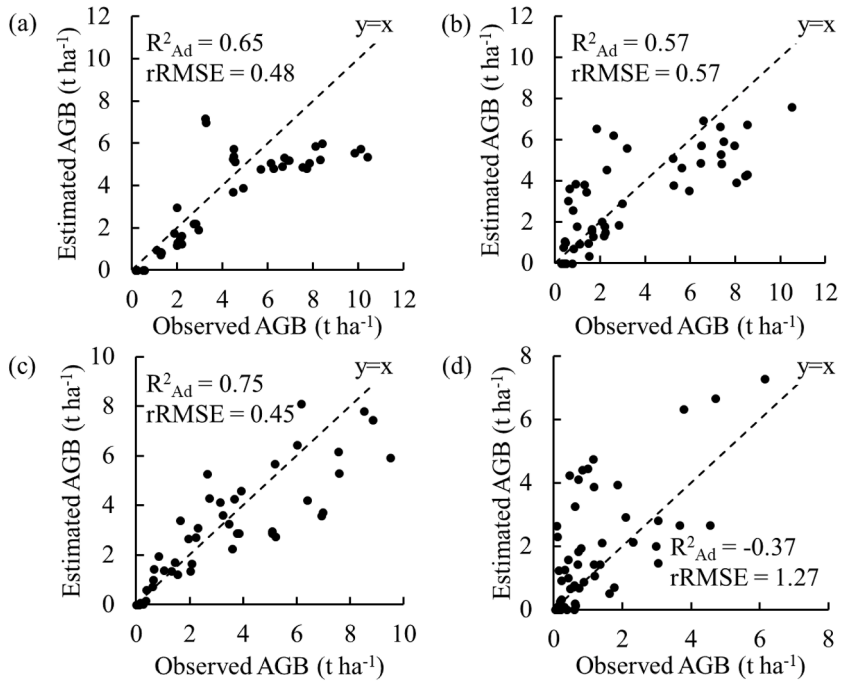

Fig. 3. Estimation results of the base model on the test dataset for four independent locations, (a) Tokyo, (b) Chiba, (c) Côte d'Ivoire, and (d) Madagascar. The dotted line means the 1:1 line.

geographically limited datasets, particularly when generalizing to new environments.

In the present study, notable overestimation for low AGB values was observed in Chiba and Madagascar, likely due to differences in soil color between the training datasets from Kyoto and test datasets from Chiba and Madagascar. While the training data from Kyoto featured grey soils, Chiba and Madagascar exhibited blackish and reddish soils, respectively, which were absent in the training data (Supplementary Fig. S10). These unseen soil types likely influenced model performance during early growth stages before canopy closure. Conversely, underestimation for high AGB values was prominent in Tokyo and Chiba but absent in Madagascar and Côte d'Ivoire (Fig. 3). This trend aligns with findings from Nakajima et al. [25]. and suggests that the lack of high AGB samples in the training dataset contributed to unstable predictions. These results highlight the importance of ensuring diverse representation in training data to account for variations in soil types and AGB distribution.

The finding that the base model trained on data from a single cultivation location did not have sufficient robustness for other locations underscores the critical need to enhance training data diversity to improve the robustness of the model. CNN models are known to rely heavily on the diversity of training data, and their generalizability diminishes when datasets are imbalanced or lack sufficient variability [37, 38]. Two complementary approaches can address this issue: collecting additional diverse training data and employing data augmentation techniques. In the present study, basic data augmentation methods, such as image flipping and brightness adjustments, were applied. While these techniques likely provided some benefit, their impact may be limited in addressing spatially dependent features. Future research should explore advanced augmentation strategies, such as synthetic data generated by Generative Adversarial Networks (GAN), to create training datasets that foster the learning of spatially invariant features. For instance, studies have demonstrated that GAN can enhance phenotyping model performance by generating diverse and realistic synthetic data [39,40].

4.3. Effect of additional data collected from a new location on the robustness

We evaluated the potential for enhancing the robustness of a CNN-based rice AGB estimation model across multiple locations. Fine-tuning is a commonly applied method for model adaptation; however, it is primarily effective when extensive and diverse training data are

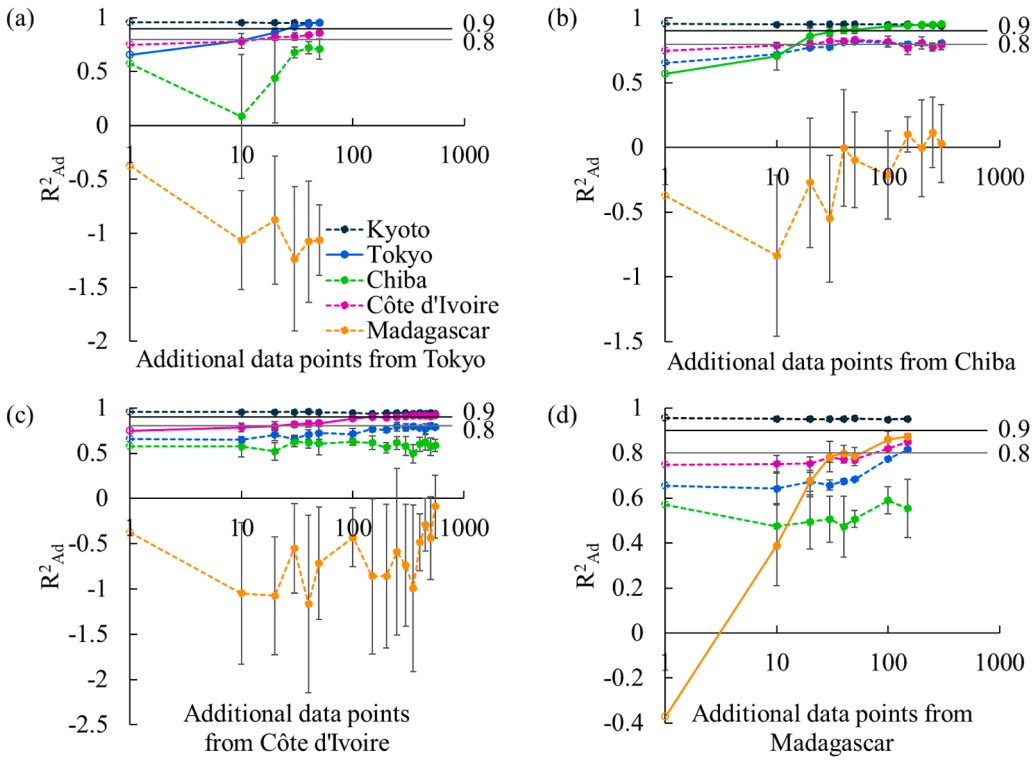

Fig. 4. Effect of adding training data from (a)Tokyo, (b) Chiba, (c) Côte d'Ivoire, and (d) Madagascar on R^2_{Ad} for each location. Solid lines represent results for the location from which data was added to the training data. Dotted lines show results for other locations. The x-axis, scaled as log (10), indicates the number of added points. The unfilled circles at $x = 1$ represent the accuracy of the base model, that is, without additional data, positioned here as $x = 0$ is undefined on a logarithmic scale.

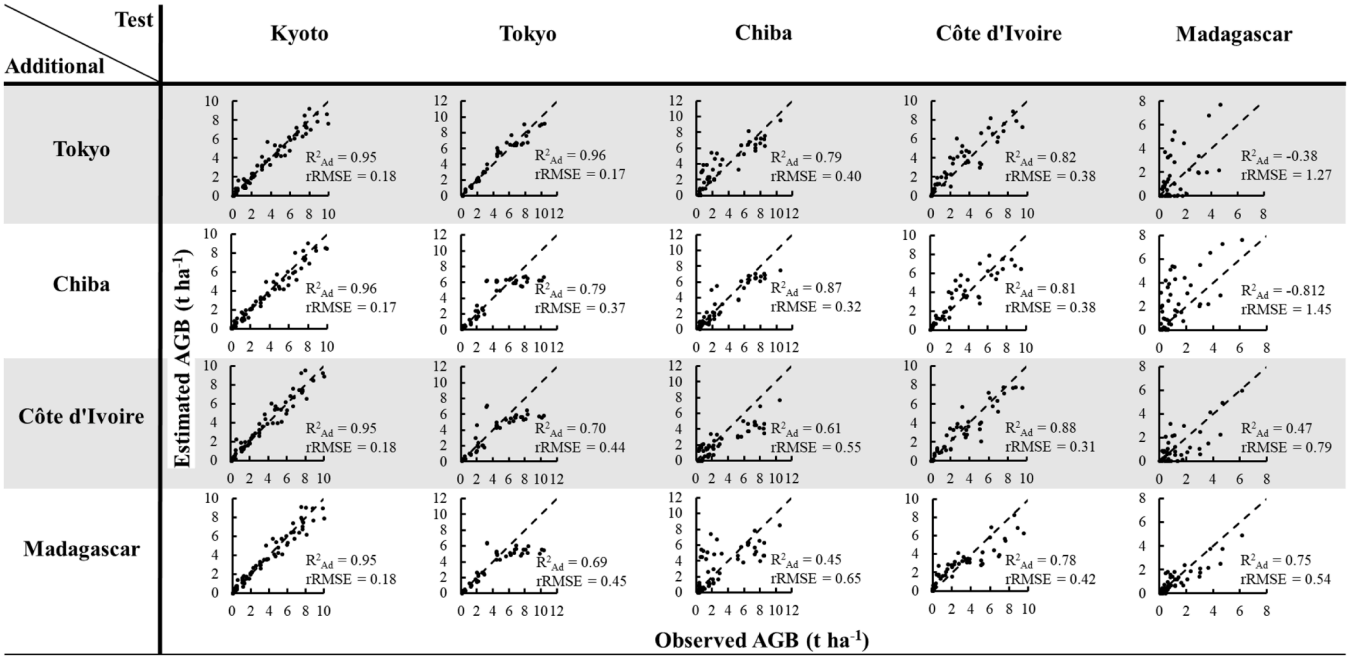

Fig. 5. Estimation results for five locations after adding 40 data points from each of four locations. Scatter plots showing estimation results are arranged in a matrix-like layout, where each row represents the location from which 40 data points were added, and each column represents the test dataset location. The estimation results for the five locations in each row are derived from the model when the lowest validation loss was recorded among the five replications. The dotted line indicates the 1:1 line.

available [41]. In AGB estimation, where data collection is costly, fine-tuning may therefore have limited efficacy. The results of the base model in the present study suggested that enhancing the diversity of training data would be crucial for establishing a general rice AGB estimation model. As a practical approach, we then tested whether adding data from an untrained location to the training dataset could improve

the model's robustness.

Adding data from locations outside Kyoto significantly improved estimation accuracy for the corresponding location (Fig. 4, Supplemental Fig. S9). Specifically, the model improved its accuracy by correcting the overestimation of lower AGB values in Madagascar and Chiba and the underestimation of higher AGB values in Tokyo and Chiba (Fig. 4, Supplemental Fig. S9). In addition, the estimation accuracy for Kyoto, where the base model was originally trained ($R^2_{Ad} = 0.96$), was consistently maintained regardless of which location's data was added. These results demonstrate that the CNN model can extract useful features from data collected across multiple locations, maintaining high accuracy within the training data's scope. To our knowledge, this study is the first to establish an AGB estimation model applicable across multiple locations from different countries.

Adding up to 40 training data points and 8 validation points (48 points in total) significantly improved the model's accuracy, even in Madagascar, where the base model's performance was notably low ($R^2_{Ad} = -0.37$ to 0.80, rRMSE=1.27 to 0.48, Fig. 3). This represents only 11 % of the base model's training data and highlights the high adaptability of CNN model with limited data addition. This feature is especially valuable for AGB estimation tasks, where data collection is costly. The additional data were selected to resemble the overall distribution of each location's dataset, ensuring a balanced representation. Interestingly, adding data from other locations also improved estimation accuracy for Tokyo and Côte d'Ivoire (Fig. 4). This implies that increasing diversity within the training dataset enables CNN model to begin capturing general features relevant to AGB estimation across locations. However, this improvement was not consistently observed for Chiba and Madagascar, suggesting that location-specific visual features might influence the model's accuracy. Furthermore, the impact of additional data varied depending on the number of data points added from a specific location. For instance, adding fewer than 100 data points from Madagascar did not substantially improve estimation accuracy for other locations (Fig. 4(d)). However, after adding >100 data points, estimation accuracy improved in Tokyo and Côte d'Ivoire, with R^2_{Ad} values exceeding 0.8. This result suggests that even data from locations with distinctive visual features, such as Madagascar, can enhance model robustness by increasing the diversity of the training dataset.

4.4. Limitations and future perspectives

The study's findings are limited by the use of data from well-managed experimental fields. Thus, the collected dataset may not reflect the visual features encountered under actual field conditions—such as severe weed infestation, pest damage, or significant lodging. Training the CNN model and evaluating its robustness with data from these diverse conditions may lead to the establishment of a truly robust and practical model for rice AGB estimation. In addition, the influence of shooting conditions on model robustness remains unclear. Rivera-Palacio et al. [42]. showed that characteristics of person who take images can significantly affect image analysis accuracy. Clarifying these effects could help separate errors due to growth conditions from those due to image acquisition, enabling more efficient model improvements.

For global assessment of AGB productivity, estimation methods must be not only robust but also highly convenient. Many conventional AGB estimation approaches rely on specialized sensors such as multispectral or hyperspectral instruments and LiDAR, which require expert knowledge and incur high costs [14]. For instance, Liu et al. [43] achieved robust potato AGB estimation across different regions by incorporating meteorological data. While including region-specific environmental data is a rational strategy to enhance model robustness, the addition of such data can impose extra constraints on the practicality of the method. In developing countries, for instance, access to meteorological information is significantly limited [44]. In contrast, the AGB estimation method established in this study relies solely on RGB images, which can be easily obtained by anyone around the world. This inherent simplicity and accessibility make the approach particularly promising for the future development of on-site, instant AGB monitoring tools.

5. Conclusion

The present study evaluated the robustness of a CNN-based rice AGB estimation model across diverse locations. While the base model trained solely on Kyoto data showed limited robustness for other untrained locations, adding up to 48 data points from each new location significantly improved accuracy, achieving practical performance ($R^2_{Ad} > 0.8$) even in challenging locations like Madagascar. Notably, adding data from one location occasionally improved the estimation accuracy at other untrained locations, highlighting the importance of data diversity for the robustness of the estimation model. Our findings not only demonstrate the importance of strategic data diversification but also provide practical insights into the minimal amount of additional data needed. These results contribute significantly toward establishing a robust, accessible, and globally applicable method for on-site monitoring of rice AGB.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (https://chatgpt.com) to improve the readability of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

We are grateful for the financial support to this study by JSPS KAKENHI grant nos. 22H02308, 24H00499, 24K01741 (to Y. Tanaka), and 24KJ1409 (to K. Nakajima). This research was also performed by the commissioned research fund provided by F-REI nos. JPFR23030101 and JPFR24030101 (to Y. Tanaka).

Ethics statement

Not applicable: This manuscript does not include human or animal research.

CRediT authorship contribution statement

Kota Nakajima: Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Kazuki Saito:** Writing – review & editing, Methodology, Investigation. **Yasuhiro Tsujimoto:** Writing – review & editing, Investigation. **Toshiyuki Takai:** Writing – review & editing, Investigation. **Atsushi Mochizuki:** Writing – review & editing, Investigation. **Tomoaki Yamaguchi:** Writing – review & editing, Investigation. **Ali Ibrahim:** Writing – review & editing, Investigation. **Salifou Goube Mairoua:** Writing – review & editing, Investigation. **Bruce Haja Andrianary:** Writing – review & editing, Investigation. **Keisuke Katsura:** Writing – review & editing, Investigation. **Yu Tanaka:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.atech.2025.100998.

Data availability

Data will be made available on request.

References

[1] D. Ray, N. Mueller, P. West, J. Foley, Yield trends are insufficient to double global crop production by 2050, Plos One 8 (6) (2013) e66428, https://doi.org/10.1371/journal.pone.0066428.

[2] X. Yin, J. Gu, M. Dingkuhn, P.C. Struik, A model-guided holistic review of exploiting natural variation of photosynthesis traits in crop improvement, J. Exp. Bot. 73 (10) (2022) 3173–3188, https://doi.org/10.1093/jxb/erac109.

[3] Z. Zhang, P. Li, L.-X. Wang, Z. Hu, L.-H. Zhu, Y. Zhu, Genetic dissection of the relationships of biomass production and partitioning with yield and yield-related traits in rice, Plant Sci. 167 (2004) 1–8, https://doi.org/10.1016/J.PLANTSCI.2004.01.007.

[4] H. Yoshida, H. Yoshida, T. Horie, K. Katsura, T. Shiraiwa, A model explaining genotypic and environmental variation in leaf area development of rice based on biomass growth and leaf N accumulation, Field Crops Res. 102 (2007) 228–238, https://doi.org/10.1016/J.FCR.2007.04.006.

[5] S. Muthayya, J. Sugimoto, S. Montgomery, G. Maberly, An overview of global rice production, supply, trade, and consumption, Ann. N. Y. Acad. Sci. 1324 (2014) 7–14, https://doi.org/10.1111/nyas.12540.

[6] G. Khush, Origin, dispersal, cultivation and variation of rice, Plant Mol. Biol. 35 (1997) 25–34, https://doi.org/10.1023/A:1005810616885.

[7] N. Nam, G. Subbarao, Y. Chauhan, C. Johansen, Importance of canopy attributes in determining dry matter accumulation of pigeonpea under contrasting moisture regimes, Crop Sci. 38 (1998) 955–961, https://doi.org/10.2135/CROPSCI1998.0011183X003800040013X.

[8] T. Shiraiwa, N. Ueno, S. Shimada, T. Horie, Correlation between yielding ability and dry matter productivity during initial seed filling stage in various soybean genotypes, Plant Prod. Sci. 7 (2004) 138–142, https://doi.org/10.1626/pps.7.138.

[9] R. Mir, M. Reynolds, F. Pinto, M. Khan, M. Bhat, High-throughput phenotyping for crop improvement in the genomics era, Plant Sci.: Int. J. Exp. Plant Biol. 282 (2019) 60–72, https://doi.org/10.1016/J.PLANTSCI.2019.01.007.

[10] L. Young, Agricultural crop forecasting for large geographical areas, Annu. Rev. Stat. Appl. 6 (2019) 173–196, https://doi.org/10.1146/ANNUREV-STATISTICS-030718-105002.

[11] L. Poley, G. McDermid, A systematic review of the factors influencing the estimation of vegetation aboveground biomass using unmanned aerial systems, Remote Sens. 12 (2020) 1052, https://doi.org/10.3390/rs12071052.

[12] A. Bazrafkan, N. Delavarpour, P. Oduor, N. Bandillo, P. Flores, An overview of using unmanned aerial system mounted sensors to measure plant above-ground biomass, Remote Sens. 15 (2023) 3543, https://doi.org/10.3390/rs15143543.

[13] J. Bendig, A. Bolten, S. Bennertz, J. Broscheit, S. Eichfuss, G. Bareth, Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging, Remote Sens. 6 (2014) 10395–10412, https://doi.org/10.3390/rs61110395.

[14] J. Ma, Y. Li, Y. Chen, K. Du, F. Zheng, L. Zhang, Z. Sun, Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network, Eur. J. Agron. 103 (2019) 117–129, https://doi.org/10.1016/J.EJA.2018.12.004.

[15] Z.Y. Wang, P. Chen, Y. Yang, H. Fu, F. Yang, M. Raza, C. Guo, C. Shu, Y. Sun, Z. Yang, Z. Chen, Estimation of rice aboveground biomass by combining canopy spectral reflectance and unmanned aerial vehicle-based red green blue imagery data, Front. Plant Sci. 13 (2022) 903643, https://doi.org/10.3389/fpls.2022.903643.

[16] F.J. Kriegler, W.A. Malila, R.F. Nalepka, W. Richardson, Preprocessing transformations and their effects on multispectral recognition, in: Proc. Sixth Int. Symp. Remote Sens. Environ., Ann Arbor, MI, USA II, 1969, p. 97, 13–16 October 1969.

[17] M. Hosseini, H. Mcnairn, S. Mitchell, L. Robertson, A. Davidson, S. Homayouni, Synthetic aperture radar and optical satellite data for estimating the biomass of corn, Int. J. Appl. Earth Obs. Geoinformation 83 (2019) 101933, https://doi.org/10.1016/J.JAG.2019.101933.

[18] Q. Jiang, S. Fang, Y. Peng, Y. Gong, R. Zhu, X.Y. Wu, B. Duan, J. Liu, UAV-based biomass estimation for rice-combining spectral, TIN-based structural and meteorological features, Remote Sens. 11 (2019) 890, https://doi.org/10.3390/rs11070890.

[19] J. Yue, G. Yang, Q. Tian, H. Feng, K. Xu, C. Zhou, Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices, ISPRS J. Photogramm. Remote Sens. 150 (2019) 226–244, https://doi.org/10.1016/J.ISPRSJPRS.2019.02.022.

[20] Y. Dai, S.T. Yu, J. Ding, K. Chen, G. Zeng, A. Xie, P. He, S. Peng, M. Zhang, Improving the estimation of rice above-ground biomass based on spatio-temporal UAV imagery and phenological stages, Front. Plant Sci. 15 (2024) 1328834, https://doi.org/10.3389/fpls.2024.1328834.

[21] H. Zheng, T. Cheng, M. Zhou, D. Li, X. Yao, Y. Tian, W. Cao, Y. Zhu, Improved estimation of rice aboveground biomass combining textural and spectral analysis of UAV imagery, Precis. Agric. 20 (2019) 611–629, https://doi.org/10.1007/s11119-018-9600-7.

[22] J. Han, L. Shi, Q. Yang, Z. Chen, J. Yu, Y. Zha, Rice yield estimation using a CNN-based image-driven data assimilation framework, Field Crops Res. 288 (2022) 108693, https://doi.org/10.1016/j.fcr.2022.108693.

[23] M. Acorsi, F. Miranda, M. Martello, D. Smaniotto, L. Sartor, Estimating biomass of black oat using UAV-based RGB imaging, Agronomy 9 (7) (2019) 344, https://doi.org/10.3390/AGRONOMY9070344.

[24] L. Wang, X. Zhou, X. Zhu, Z. Dong, W. Guo, Estimation of biomass in wheat using random forest regression algorithm and remote sensing data, Crop J. 4 (2016) 212–219, https://doi.org/10.1016/J.CJ.2016.01.008.

[25] K. Nakajima, Y. Tanaka, K. Katsura, T. Yamaguchi, T. Watanabe, T. Shiraiwa, Biomass estimation of World rice (Oryza sativa L.) core collection based on the convolutional neural network and digital images of canopy, Plant Prod. Sci. 26 (2023) 187–196, https://doi.org/10.1080/1343943X.2023.2210767.

[26] Z. Lu, H. Xie, C. Liu, Y. Zhang, Bridging the gap between vision transformers and convolutional neural networks on small datasets, ArXiv (2022), https://doi.org/10.48550/arXiv.2210.05958.

[27] A. Sasaki, M. Ashikari, M. Ueguchi-Tanaka, H. Itoh, A. Nishimura, D. Swapan, K. Ishiyama, T. Saito, M. Kobayashi, G. Khush, H. Kitano, M. Matsuoka, Green revolution: a mutant gibberellin-synthesis gene in rice, Nature 416 (2002) 701–702, https://doi.org/10.1038/416701a.

[28] B. Yu, Z. Lin, H. Li, X. Li, J. Li, Y. Wang, X. Zhang, Z. Zhu, W. Zhai, X. Wang, D. Xie, C. Sun, TAC1, a major quantitative trait locus controlling tiller angle in rice, Plant J.: Cell Mol. Biol. 52 (5) (2007) 891–898, https://doi.org/10.1111/J.1365-313X.2007.03284.X.

[29] Y. Kojima, K. Ebana, S. Fukuoka, T. Nagamine, M. Kawase, Development of an RFLP-based Rice diversity research set of germplasm, Breed. Sci. 55 (2005) 431–440, https://doi.org/10.1270/JSBBS.55.431.

[30] Y. Tanaka, T. Watanabe, K. Katsura, Y. Tsujimoto, T. Takai, T. Tanaka, K. Kawamura, H. Saito, K. Homma, S. Mairoua, K. Ahouanton, A. Ibrahim, K. Senthilkumar, V. Semwal, E. Matute, E. Corredor, R. El-Namaky, N. Manigbas, E. Quilang, Y. Iwahashi, K. Nakajima, E. Takeuchi, K. Saito, Deep learning enables instant and versatile estimation of rice yield using ground-based RGB images, Plant Phenomics 5 (2023) 0073, https://doi.org/10.34133/plantphenomics.0073.

[31] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, pp. 770–778, https://doi.org/10.1109/cvpr.2016.90.

[32] G. Huang, Z. Liu, K. Weinberger, Densely connected convolutional networks, in: Proc. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.

[33] D. Kingma, J. Ba, Adam: a method for stochastic optimization, ArXiv (2014), https://doi.org/10.48550/arXiv.1412.6980.

[34] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zhang, TensorFlow: a system for large-scale machine learning, ArXiv (2016), https://doi.org/10.48550/arXiv.1605.08695.

[35] L.N. Habibi, T. Matsui, T. Tanaka, Critical evaluation of the effects of a cross-validation strategy and machine learning optimization on the prediction accuracy and transferability of a soybean yield prediction model using UAV-based remote sensing, J. Agric. Food Res. 16 (2024) 101096, https://doi.org/10.1016/j.jafr.2024.101096.

[36] M. Ferraciolli, F. Bocca, L. Rodrigues, Neglecting spatial autocorrelation causes underestimation of the error of sugarcane yield models, Comput. Electron. Agric. 161 (2019) 233–240, https://doi.org/10.1016/J.COMPAG.2018.09.003.

[37] K. Zhou, Y. Yang, Y. Qiao, T. Xiang, Domain generalization with MixStyle, ArXiv (2021), https://doi.org/10.48550/arXiv.2104.02008.

[38] D. Dablain, K. Jacobson, C. Bellinger, M. Roberts, N. Chawla, Understanding CNN fragility when learning with imbalanced data, ArXiv (2022), https://doi.org/10.48550/arXiv.2210.09465.

[39] Z. Hartley, A. French, Domain adaptation of synthetic images for wheat head detection, Plants 10 (12) (2021) 2633, https://doi.org/10.3390/plants10122633.

[40] Y. Li, X. Zhan, S. Liu, H. Lu, R. Jiang, W. Guo, S. Chapman, Y. Ge, B. Solan, Y. Ding, F. Baret, Self-supervised plant phenotyping by combining domain adaptation with 3D plant model simulations: application to wheat leaf counting at seedling stage, Plant Phenomics 5 (2023) 0041, https://doi.org/10.34133/plantphenomics.0041.

[41] M. Welch, C. McIntosh, A. Traverso, L. Wee, T. Purdie, A. Dekker, B. Haibe-Kains, D. Jaffray, External validation and transfer learning of convolutional neural networks for computed tomography dental artifact classification, Phys. Med. Biol. 65 (2020) 035017, https://doi.org/10.1088/1361-6560/ab63ba.

[42] J.C. Rivera-Palacio, C. Bunn, M. Ryo, Factors affecting deep learning model performance in citizen science–based image data collection for agriculture: a case study on coffee crops, Comput. Electron. Agric. 232 (2025) 110096, https://doi.org/10.1016/j.compag.2025.110096.

[43] Y. Liu, Y. Fan, J. Yue, X. Jin, Y. Ma, R. Chen, M. Bian, G. Yang, H. Feng, A model suitable for estimating above-ground biomass of potatoes at different regional levels, Comput. Electron. Agric. 222 (2024) 109081, https://doi.org/10.1016/j.compag.2024.109081.

[44] L. Georgeson, M. Maslin, M. Poessinouw, Global disparity in the supply of commercial weather and climate information services, Sci. Adv. 3 (5) (2017) e1602632, https://doi.org/10.1126/sciadv.1602632.