

# Deep learning-based estimation of rice yield using RGB image

**Yu Tanaka** (✉ [tanaka.yu.2s@kyoto-u.ac.jp](mailto:tanaka.yu.2s@kyoto-u.ac.jp))

Kyoto University <https://orcid.org/0000-0003-2106-4217>

**Tomoya Watanabe**

Kyushu University

**Keisuke Katsura**

Tokyo University of Agriculture and Technology

**Yasuhiro Tsujimoto**

Japan International Research Center for Agricultural Sciences

**Toshiyuki Takai**

NARO Institute of Crop Science

**Takashi Tanaka**

Gifu University <https://orcid.org/0000-0001-7116-6962>

**Kensuke Kawamura**

Japan International Research Center for Agricultural Sciences

**Hiroki Saito**

Japan International Research Center for Agricultural Sciences

**Koki Homma**

Tohoku University

**Salifou Mairoua**

Africa Rice Center (AfricaRice)

**Kokou Ahouanton**

Africa Rice Center (AfricaRice)

**Ali Ibrahim**

Africa Rice Center (AfricaRice), Regional Station for the Sahel

**Kalimuthu Senthilkumar**

Africa Rice Center (AfricaRice)

**Vimal Semwal**

Africa Rice Center (AfricaRice), Nigeria Station

**Eduardo Matute**

Latin American Fund for Irrigated Rice

**Edgar Corredor**

Latin American Fund for Irrigated Rice

**Raafat El-Namaky**

Rice Research and Training Center <https://orcid.org/0000-0001-9671-8954>

**Norvie Manigbas**

Philippine Rice Research Institute (PhilRice)

**Eduardo Quilang**

Philippine Rice Research Institute (PhilRice)

**Yu Iwahashi**

Kyoto University <https://orcid.org/0000-0002-1291-6329>

**Kota Nakajima**

Kyoto University

**Eisuke Takeuchi**

Kyoto University

**Kazuki Saito**

Africa Rice <https://orcid.org/0000-0002-8609-2713>

---

**Article**

**Keywords:** deep learning, crop yield, rice, convolutional neural network

**Posted Date:** October 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1026695/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Crop productivity is poorly assessed globally. Here, we provide a deep learning-based approach for estimating rice yield using RGB images. During ripening stage and at harvest, over 22,000 digital images were captured vertically downwards over the rice canopy from a distance of 0.8 to 0.9 m, and rice yields were obtained in the corresponding area ranging from 0.1 and 16.1 t ha<sup>-1</sup>. A convolutional neural network (CNN) applied to these data at harvest predicted 70% variation in rice yield with a relative root mean square error (rRMSE) of 0.22. Images obtained during the ripening stage can also be used to forecast the final rice yield. Our work suggests that this low-cost, hands-on, and rapid approach can provide a breakthrough solution to assess the impact of productivity-enhancing interventions and identify fields where these are needed to sustainably increase crop production.

## Introduction

The global demand for staple crop products is expected to increase by 60% by 2050, mainly because of the increased population, per capita income growth, and use of biofuels<sup>1</sup>. To meet this estimated future demand, crop production must be enhanced in an environmentally sustainable manner in the context of increasing competition for water, land, and labour, and under potentially more extreme weather conditions associated with climate change<sup>2</sup>. As conversion of carbon-rich and biodiverse natural ecosystems to cropland causes greenhouse gas emissions and further climate change, it is necessary to make effective use of the existing cropland to further increase production through sustainable intensification to increase yield and reduce yield gap, while reducing negative environmental impacts<sup>3,4</sup>. Furthermore, agriculture needs to address problems of poverty, poor food, and nutrition security for smallholder farmers. Despite the importance of these goals in agriculture, crop productivity is poorly assessed, especially in the global South, where there is need to monitor agricultural productivity and evaluate the impact of productivity-enhancing interventions<sup>5</sup>. There are three well-known approaches for assessing crop yield, which include self-reporting, crop cutting, and remote sensing technologies. However, self-reported data from smallholder farmers are often inaccurate<sup>9</sup>. Crop cut, wherein a sub-section of a plot is physically harvested, is time- and labour-consuming, and difficult to scale to large areas with financial limitations. Remote sensing technologies require expensive instruments such as satellites, unmanned aerial vehicles (UAVs), and specialised sensors in many cases, which makes them difficult for practical use in the global South. The absence of reliable data on agriculture statistics is a serious constraint for both agricultural research and policy.

With recent advancement in computational technology, ground-based images captured by low-cost devices together with so called “machine learning” approaches have received great interest. Machine learning technology is one of the most remarkable innovations in the last decade<sup>7,8</sup>. Deep learning is categorised as supervised machine learning and mainly consists of convolutional neural networks (CNNs). A remarkable feature of CNN is its capability for image analysis. It has already been applied in various situations, which include language translation<sup>9</sup>, protein structure prediction<sup>10</sup>, board games<sup>11</sup>,

and agriculture. To develop a practical CNN model, a large-scale combination of images and supervising data is required. The desirable target objects or crop characteristics could be those that are relatively easy to be visually evaluated for massive data collection. For these reasons, many earlier studies applying CNNs to agriculture focused on the classification of crop biotic<sup>12,13,14</sup> and abiotic stresses<sup>15</sup>, and estimation of crop growth-related traits such as biomass<sup>16,17,18,19</sup>, leaf area index<sup>20</sup>, grain number<sup>21</sup>, and panicle density<sup>22,23</sup>, which could help indirectly predict crop yield through the use of crop simulation models and their empirical relationships with yield. However, to the best of our knowledge, no study has directly estimated crop yield using deep learning with ground-based images.

This study focuses on rice, which is by far the most important in terms of human consumption in low- and lower-middle income countries among the big three cereals and is mainly cultivated by smallholder farmers<sup>24</sup>. We established a database of ground-based digital images of rice taken during the ripening stage and at harvest, and the corresponding yields were collected from seven countries using a standardised data collection procedure. We then developed a CNN model that covered a wide range of yield levels, rice growing environments, cultivars, and crop management practices, such as crop establishment methods and fertiliser management. We assessed the robustness of the model under various conditions which potentially affected the yield estimation. We demonstrate that rice yield can be rapidly and effectively estimated at a low cost without involving labour-intensive crop cuts or expensive remote-sensing technologies at harvest and during the ripening stage with satisfactory accuracy.

## Results

### Database on rice canopy image and grain yield

The multinational dataset of rice canopy image and corresponding rough and filled grain yields, and aboveground dry weight was established with a standardised data collection procedure for 4820 harvested plots and 22067 images in various on-station and on-farm field experiments and farmers' fields or seed production plots across 20 locations in seven countries (Fig. 1a, Supplementary Fig S1, Supplementary Table S1, S3). The database includes 415 plots from on-farm fields accounting for 9% of the total plots. Cote d'Ivoire, Senegal, and Japan accounted for 56%, 32%, and 5% of total data points, respectively (Fig. 1b). The dataset covers both lowland and upland rice production systems containing 462 rice cultivars, and include two crop establishment methods (direct seeding and transplanting) (Supplementary Table S2). N-P-K fertiliser application ranged from 0 to 200 kg N ha<sup>-1</sup>, 0 to 120 kg P<sub>2</sub>O<sub>5</sub> ha<sup>-1</sup>, and 0 to 120 K<sub>2</sub>O kg ha<sup>-1</sup>, respectively (Supplementary Table S1). The observed rough grain yield ranged from 0.1 to 16.1 t ha<sup>-1</sup> with an average of 5.8 t ha<sup>-1</sup> and showed a normal distribution (Fig. 1a). Hence, our dataset covers a wide range of yield levels, crop management, cultivars, and growing environments of rice (Supplementary Table S1). We found strong and positive relationships between rough grain yield, aboveground dry weight, and filled grain yield (Supplementary Fig. S2). Further data analyses using the CNN model focused only on rough grain yield. The main part of the dataset was split into three parts: (i) development and evaluation consisting of training (72% of harvested plots in this

development and evaluation), validation (14%), and test (14%); (ii) robustness; and (iii) prediction (Fig. 1c). The prediction dataset consisted of data collected in Moshi, Tanzania, and in Tokyo, Japan. This implies that the prediction accuracy of the developed CNN model was evaluated at the “unknown” and “independent” dataset in this study. Furthermore, among the five cultivars grown in these locations, one cultivar in Tanzania (cv. TXD 306) was not included in any other dataset.

## A CNN model to estimate rough grain yield from canopy image

The CNN structure used in the present study has five convolutional layers with one fully connected layer in the main stream with three branching layers (Supplementary Fig S3). The learning rate and batch size during the learning process were optimised with 10 replications. The combination of learning rate and batch size of 0.0001 and 32, respectively, resulted in the best performance for the test dataset (Supplementary Figure S4). With this combination, the best model of the learning process was generated at epoch = 61, and the model was used for all of the following analyses (Fig. 2a). The developed CNN model could explain approximately 70% of the variation in yield for validation and test data, respectively, with a relative root mean square error (rRMSE) of approximately 0.22 for both (Fig. 2b-c). The relationship between the observed and estimated yields fit well to the 1:1 line for both datasets. The deviation between the estimated and observed yields of individual cultivars in the test dataset was plotted against the number of harvested plots in the training dataset (Fig. 2d). The cultivars with more than 25 plots in the training dataset tended to have less than 1 t ha<sup>-1</sup> deviation. The empirical relationships illustrated as upper and lower boundary curves in Fig. 2d indicate that increasing the number of data points by 10 times can reduce the error of the yield estimation by 50%.

The accuracy of the CNN model was further evaluated using the prediction dataset. The model estimated the rough grain yield with an R<sup>2</sup> of 0.487 and rRMSE of 0.174 across five cultivars in two countries (Fig. 3a). It underestimated the yield of the cv. Koshihikari; the deviation between the estimated and observed yield was higher in this cultivar than in the other cultivars. However, the model successfully estimated the yield variation observed in the cv. TXD 306, which was included solely in the prediction dataset. To determine the number of images that should be used for proper yield prediction per plot, rough grain yield was predicted using different numbers of replicated images (1–5) per harvested plot (Fig. 3b), and averaged across images per harvest plot. There were no apparent differences in R<sup>2</sup> and rRMSE between the observed and predicted yields using different numbers of images, with R<sup>2</sup> of 0.469 to 0.491 and rRMSE of 0.175 to 0.180. When comparisons were made among the predicted yields using different numbers of images, they were strongly and positively correlated.

To understand how the CNN model reads the images and estimates rice yield, we used the occlusion-based visualisation technique to estimate the additive effect on yield estimation<sup>25</sup>. Briefly, the specific part of the image was masked by a grey square, and the yield estimation of the masked image was subtracted from that of the original image. The calculated values can be interpreted as the additive effect

of the masked region on the yield estimation and mapped to the original image with a colour scale (Supplementary Fig S5). This analysis revealed that the regions containing many rice panicles have a positive effect, whereas the region with leaves, stems, or ground has a negative effect on yield estimation. The importance of the panicles for yield estimation was further validated using panicle removal experiments conducted in Kyoto, Japan. The two panicles per hill were sequentially removed from the canopy, and the rough grain weight and canopy images were recorded for each sequence (Fig. 4a-b). The yield was estimated using the CNN model for each sequence of panicle removal. The heat map analysis confirmed that the regions containing many panicles had a positive effect on yield estimation in the initial rice canopy. However, these regions diminished as the panicles were gradually removed (Fig. 4c). However, when panicles were removed, the regions with overlapping or senescent leaves in the lower position tended to have positive effects (Fig. 4d). The estimated yield for the canopy with no panicles was  $1.60 \text{ t ha}^{-1}$ , which implied that apart from the existence of panicles, information on the background canopy may have also been utilised for yield estimation.

## Robustness of the developed CNN model

The robustness of the CNN model to image quality was tested using the images taken (i) from different shooting angles, (ii) at various times of day during the five days before harvest, and (iii) on different shooting dates during the ripening stage. The shooting angle assumes human error, while the time of day reflects the changing natural environment causing the variation of the contrast or colour balance of the image. The shooting date is important to assess when rice yield can be effectively predicted by using our model during the ripening stage.

To determine the range of shooting angles acceptable for the developed CNN model, we estimated rice yield using images acquired from eight shooting angles (in  $10^\circ$  increments from  $20^\circ$  to  $90^\circ$  (control)) in Mbe, Côte d'Ivoire (Fig. 5a). The deviation between the estimated and observed yields was averaged across 25 harvested plots at each angle. The deviation ranged from  $-3.7$  to  $2.4 \text{ t ha}^{-1}$  when the depression angle was  $20^\circ$  (Fig. 5b). The deviation decreased with an increase in the depression angle. When the outlier was excluded, the ranges of the deviation were between  $-0.45$  and  $2.44 \text{ t ha}^{-1}$  at  $60^\circ$ , which was comparable with that at  $90^\circ$  (control). The heat map analysis with images taken at a shallower angle showed that the regions having an inner structure of the canopy, such as stems or leaves in the lower position tended to have a significant negative effect on the yield estimation (Fig. 5c). Furthermore, the regions with overlap between the leaves and panicles in images at shallower angles, such as  $20^\circ$  to  $50^\circ$  did not have a positive effect like the image in the control angle (e.g., left bottom parts in the  $20^\circ$  image, and upper parts in the  $50^\circ$  image). The estimation accuracy analysis showed that greater depression angles resulted in better estimation accuracy (Fig. 5d). When the depression angle was greater than  $60^\circ$ , the  $R^2$  and rRMSE calculated between the estimated and observed yields ranged from 0.435 to 0.493 and 0.180 to 0.219, respectively. Strong correlations were found among the estimated yields from  $70^\circ$ ,  $80^\circ$ , and  $90^\circ$ , with  $R^2$  greater than 0.76 and rRMSE of less than 0.11.

The image of the rice canopy was captured by a fixed-point camera every 30 min for 5 successive days before and at harvest in Kyoto, Japan (Fig. 6a). The images for every 2 h on 29 August 2020 are shown as an example of a clear sunny day (Fig. 6c). The image taken at 0600 hrs has a different colour balance compared to the others because of the lower irradiation. The images taken at 0800 hrs, 1400 hrs, and 1600 hrs have higher contrast because of the shallower angle of solar radiation. The images taken at 1000 hrs and 1200 hrs were bright and had lower contrast because of the greater angle and intensity of the solar radiation. Despite such variation in light environments, the CNN model provided stable outputs throughout the daytime with a slight overestimation (Fig. 6b). The heat map analysis revealed that the CNN model showed stable recognition of the panicles regardless of the image quality (Fig. 6d), which led to a robust estimation of yield.

To assess from when the CNN model can predict rice yield during ripening stage, the canopy image was taken once a week after 50% heading until the harvest for 22 cultivars in Mbe, Cote d'Ivoire. The yield estimated in the early ripening stage tended to have a lower yield than the observed yield at harvest, whereas such a trend was not observed with the yield estimated at the later ripening stage (Fig. 7a). This indicates that the model recognises mature panicles (Fig. 4, Supplementary Fig S5) but not the immature panicles. When the data from 22 cultivars was pooled, the ratio of the estimated yield to the observed yield ranged from 0.3 to 0.6 at just after 50% heading, and the y-intercept of the segmented regression was 0.517. The ratio increased linearly during ripening. The relationship reached a plateau at approximately 4 weeks after 50% heading (WAH) (Fig. 7b). A similar trend was also observed in Madagascar (Supplementary Fig. S6), while the relative yield plateaued within 2 WAH. The  $R^2$  values between the estimated yield during 2 to 4 WAH and the observed yield ranged from 0.370 to 0.410, whereas it was 0.572 at harvest (Fig. 7c). The rRMSE between the estimated yield after 3 WAH and the observed yield ranged from 0.193 to 0.196. When comparisons were made among the estimations after 3 WAH, the  $R^2$  and rRMSE ranged from 0.657 to 0.767 and 0.135 to 0.228, respectively. The CNN model slightly underestimated the yield at 3 WAH compared with the observed yield, whereas it slightly overestimated the yield at harvest (Fig. 7d). The correction of the estimated yield by using the empirical relationship observed in Fig. 7b was conducted to reduce the deviation, especially in the earlier ripening stage. When the corrected estimation at 2 WAH was compared with the observed yield,  $R^2$  and rRMSE improved to 0.381 and 0.196, respectively (Fig S7).

## Discussion

Multinational database of canopy image at harvest and rough grain yield collected using the standardised data collection procedure in a wide range of rice growing conditions (Fig. 1, Supplementary Fig S1, S2) contained more than 22,000 images, and had large variation in rice yield. This dataset enabled the development of a CNN model that can estimate rice yield under a wide range of conditions (Fig. 2a, Supplementary Fig S3, S4). No other studies have developed a model to predict rice yield accurately only by using RGB images captured with a commercially available digital camera. The results from our analysis using the prediction dataset, which are independent of the others and include the

unique cultivar cv. TXD 306, clearly demonstrate that our CNN model is capable of estimating rice grain yield.

It is repeatedly reported that satellite data alone or in combination with other data and models can estimate crop growth-related traits such as aboveground biomass and leaf area index, and indirectly predict crop yield in farmers' fields<sup>26,27,28,29</sup>. UAVs were proposed as a powerful tool for estimating the aboveground biomass by utilising various sensors<sup>30,31</sup>. The accuracy of estimation directly using rice canopy images in the present study is comparable to or even higher than those shown in earlier studies. The accuracy of our model was achieved without any expensive equipment. Furthermore, the accuracy was evaluated using an independent prediction dataset, which has rarely been tested in earlier studies. Our model was able to estimate rice yield with satisfactory prediction in the existing most comprehensive dataset in terms of the growing environments, camera settings, and number of cultivars. The object detection algorithm based on CNN enabled the detection of rice<sup>22</sup> and wheat<sup>23</sup> panicles, and it can be a potential approach for indirect yield estimation. However, it is well known that other yield components interact with panicles and strongly affect rice yield<sup>32</sup>. Unless the models for predicting other yield components are not developed, the model for detecting panicles would not be useful for yield estimation.

However, the unknown conditions causing the poor estimations of the CNN model should always be assumed when considering the scale and diversity of rice cropping systems globally. For instance, the dataset does not include the canopy affected by severe lodging, pests, insects, weeds, or abiotic stresses such as heat, drought, and flooding. Most of the data points are from irrigated lowland rice fields with relatively higher yields, and data from farmers' fields are limited. Thus, further data collection is required, especially for low-yielding and rainfed environments, and assessment of the potential use of the model for stressed or injured rice plants is warranted. The most practical solution to adopting the model to these new conditions would be to add these new data to the database and develop a new model. The results in Fig. 2d suggest that better accuracy can be achieved with more data points. As a criterion, approximately 25 harvesting plots are needed for adaptation to new conditions with practical accuracy, which should be validated for developing a sampling framework for improving and adapting the model to new conditions.

The occlusion-based method for visualising the distribution of the additive effect on the yield estimation clearly indicates that the CNN model autonomously learned the contribution of panicles on yield only by the relationships between input canopy images and the observed yield (Fig. 4c, Fig. 5c, 6d, and Supplementary Fig S5). However, the CNN model predicted a positive value of yield for canopies with no panicles in the panicle removal experiment (Fig. 4d). Similarly, the model estimated the positive value of the yield for the images taken at around 50% heading date when the panicles were immature (Fig. 7b, Supplementary Fig S6). Although the canopy used in the panicle cut experiment is unrealistic as it has substantial biomass without panicles at harvest (Fig. 4b), these results imply that the model could have also utilised the information on background canopy, such as the amount of leaves, planting density, or stem size for yield estimation.

The robustness of the CNN model to image quality is crucial because the image is not necessarily acquired under optimal rice growing conditions. Based on our assessment of the robustness of the model, the results suggest that (i) the model can be applied to the depression angles of the camera from 60° to 120° (Fig. 5), (ii) the model output is slightly affected by the changing light intensity without any reference board or colour checker (Fig. 6b), and (iii) forecasting the yield prior to the harvest is possible using the model and images acquired at 3 WAH or later. Three WAH corresponded to approximately 10 to 20 days before harvest (Supplementary Fig S8). We also found that a single image per plot was sufficient for a proper estimation of yield. These results clearly show that the CNN model offers great advantages for application in field conditions. Particularly, yield forecasting has great potential benefits in terms of field management, marketing, distribution, and policy decisions. By correcting the output of the CNN model based on the relationship shown in Fig. 7b, the yield may be forecasted even earlier than 3 WAH (Supplementary Fig S7). However, this relationship seems to be different across growing conditions and set of cultivars, and the ratio of estimated to observed yield saturated earlier in Madagascar than in Cote d'Ivoire (Fig. 7b, Supplementary Fig S6). The reason for such differences between the two locations is not known, although it may be a combined effect of various factors such as cultivar-specific dynamics of grain-filling, growing environment, soil fertility, and water management, and therefore, further studies are warranted. Additionally, the robustness of the CNN model can be evaluated at different distances from the canopy, which further enhances the applicability of the model in the future.

The CNN structure used in this study has several convolutional layers (Supplementary Fig S3), and is much smaller than the representative structures for image recognition<sup>33</sup>. This implies that the developed model can be easily transferred to mobile devices such as smartphones. The model does not require any type of colour checker. It can accept the depression angle of the image from 60° to 120° at any time of the day, at 3 WAH or later, for shooting the canopy image. The flexibility and robustness of the developed model provide a breakthrough solution for non-destructive, rapid, and on-site evaluation of rice productivity, which enables the assessment of the impact of productivity-enhancing interventions and identifying fields where these are needed to sustainably increase crop production.

## Methods

### *Construction of database for rice canopy image and rough grain yield.*

Field campaigns were conducted in 2019 and 2020 at 20 locations in seven countries (Côte d'Ivoire, Senegal, Japan, Kenya, Madagascar, Nigeria, and Tanzania). Data on rice growth traits and digital images were collected in seed production plots as well as experimental fields at research stations and farmers' fields (Supplementary Table S1). At maturity, the RGB images were captured vertically downwards over the rice canopy from a distance of 0.8 to 0.9 m using a digital camera (Fig S1a). The digital cameras used in this study are listed in Table S1. Five images were taken per harvesting plot by slightly shifting the camera for image augmentation. The rice canopy images cover approximately 1 m<sup>2</sup>, which correspond to the harvesting area proposed by Food and Agriculture Organisation (FAO) and used by Japan for agricultural statistics<sup>34</sup>. Rough grain yield that contained filled and unfilled grains was

measured at the corresponding plot or larger plots, where yield data were collected based on field experiments (Supplementary Table S1). Rice yields were reported as 14% moisture. The aboveground total dry weight and filled grain weight were also recorded in most studies. Rice yield level, rice production system, rice variety, and key crop management practices are shown in Supplementary Table S1. The database consists of eight categories, as presented in Fig 1c. For most of the training, validation, and test data, we used only a single image per plot. These three categories are the main part of the database and randomly split by a ratio of approximately 72:14:14. After splitting the data, the images were augmented for 4-fold by flipping horizontally, vertically, and their combination, which resulted in 17764 images for training data. For panicle removal, angle, shooting date (see the following sections), and prediction data, we used five replicated images per plot. The prediction data consisted of the dataset collected at Moshi (3.45S, 37.38E), Tanzania, and at Tokyo (35.41N, 139.29E), Japan, where the data were not included in any other categories. For the time-of-day data, the sequential shooting of the canopy images was conducted using a fixed camera. In total, 4820 yield data and 22067 images of 462 rice cultivars were used in this study (Figure 1c, Supplementary Table S2).

#### *Panicle removal, and experiments for robustness evaluation*

The panicle removal experiment was conducted at Kyoto (35.2N, 135.47E) and Tsukuba (36.03N, 140.04E), Japan. The five replicated canopy images were acquired for the plot to be harvested. Two panicles per hill at the random position of the canopy was removed, and then 5 images were acquired. The grain weight from the collected panicles were measured separately. By repeating this process until all the panicles were removed from the harvesting plot, the series of images with gradually decreased panicle number and the corresponding yield were obtained. The dataset at Tsukuba was included for the training, validation, and test data, and the dataset at Kyoto was used to evaluate the impact of canopy removal on the yield estimation.

The angle changing experiment was conducted at M'bé (7.87N, 5.11W), Cote d'Ivoire. The curved rail with a diameter of 1.8 m was fixed above the canopy to be harvested. By shifting the position of the camera on the rail, the image from the various depression angles were shot with the constant centre of the image. The depression angles were set to 20, 30, 40, 50, 60, 70, 80, and 90 (control) degrees. The data for angle changing experiment was collected for 25 harvested plots. The day time experiment was conducted at Kyoto, Japan. HykeCam SP2 (Hyke Inc., Japan) was fixed above the canopy of cv. Koshihikari and Takanari. The canopy images were automatically recorded every 30 min 5 days before the date of harvest for Koshihikari, and 11 days prior to harvest for Takanari. After finishing the record, the plot was harvested by the common protocol with other experiments. The data of Takanari was used for the model development and the data of Koshihikari was used for the time-of-day analysis.

The shooting date experiment was conducted at M'bé, Cote d'Ivoire and Marovoay, Madagascar. At M'bé, the 22 cultivars grown in 34 plots in total were used. The canopy images of these plots were acquired once a week from 1 to 4 weeks after 50% heading, 2 days and 1 day before harvest, and at harvest. Only

the images taken on 2 days and 1 day before harvest were used for model development, while the others were used for the shooting date analysis. After the final image records, the rice plants were harvested using a common protocol. At Marovoay, the canopy images of seven plots were recorded from 2 days prior to 14 days after 50% heading. Six images were taken every 10 min from 1200 to 1250 hrs and were used for the shooting date analysis.

### *Image processing and development of convolutional neural network model*

The RGB images of the rice canopy were recorded with an aspect ratio of 4:3 or 16:9. For the images recorded at 16:9, the edge of the long side was trimmed to a ratio of 4:3. The images were then resized to  $450 \times 600$  pixels for recording in the database, and again resized to a square of  $512 \times 512$  pixels in 8-bit PNG format as inputs for the CNN model. A bilinear algorithm was used to resize the images. The brightness values of each channel of RGB were divided by 255 to scale from 0 to 1. These values were then standardised using the mean and variance calculated from all images categorised in the training dataset. The mean and variance of the RGB channel for the training dataset were  $[R, G, B] = [0.490, 0.488, 0.281]$  and  $[0.230, 0.232, 0.182]$ , respectively. The structure of the CNN was determined using an automated structure search by Neural Network Console software (Sony Network Communications Inc., Japan). The determined CNN structure (Supplementary Fig S3) was then deployed using Python language (version 3.7) with Pytorch framework (version 1.7). The loss function and optimizer were defined by the mean absolute error and Adam optimizer, respectively. The optimal learning rate and batch size were determined by changing the combination of these hyper-parameters. Batch sizes of 16, 32, 64, 128, and learning rates of 0.0001, 0.0002, 0.0005, 0.0008, and 0.001 were combined, and the learning process was replicated 10 times for each combination. The epoch number was set to 100, and the learning process was conducted by minimising the loss of estimated and observed yields in the training dataset. The validation loss was also calculated for every epoch, and the model showing the least loss for validation was recorded. The rRMSE for the test dataset was calculated for models with all combinations of the hyper-parameters, and averaged across 10 replications. The best combination of batch size and learning rate was determined, and the recorded model was used in the present study.

### *Occlusion-based method to quantify the additive effect on the yield estimation*

The occlusion-based method<sup>25</sup> was applied to visualise the spatial distribution of the additive effect on yield estimation. The image of the rice canopy with  $450 \times 600$  pixels was partly masked by the grey square with a brightness of  $[R, G, B] = [0.5, 0.5, 0.5]$ . The size of the grey square was  $30 \times 30$  pixels. By shifting the position of the grey square by 30 pixels for both the row and column directions of the image array, 300 images were generated per original image (Supplementary Fig. S5a, b). Each portion of the original image was covered by one of the images in a series of 300 images with a grey square. Then, the rough grain yield was estimated using the CNN model, and the subtraction against the estimation for the

original image was calculated. These values overlapped with the original image as a heat map (Supplementary Fig S5c).

### *Statistical analyses, data summarizing, and code availability*

The 4820 observations of rough grain yield data were summarised by calculating the average, maximum, and minimum yields. The data were categorised according to the collected country, and the average yield in each country was calculated. The  $R^2$  and rRMSE were calculated to evaluate the model performance in each analysis. The rRMSE is defined as follows:

$$\frac{1}{\bar{y}} \sqrt{\frac{1}{n} \sum_{k=1}^n (f_i - y_i)^2} \quad (1)$$

where  $\bar{y}$ , is the average of the observed yield,  $n$  is the size of the data, and  $f_i$  and  $y_i$  are the individual estimations and observations of the yield. The rough grain yield for panicle removal, angle, shooting date, and prediction dataset was estimated with five replicated images per harvested plot, and then averaged. The standard error of the five replicated estimations was calculated in the panicle removal experiment. For the changing angle experiment, the first, second, and third quartiles were calculated for the deviation between the estimated and observed yields across 25 plots and displayed with their average, maximum, and minimum values as the box plot. For the day time experiment, the estimated yield for every 30 min was averaged across successive 6 days, and the standard error was calculated. Segmented linear regression was adopted to determine the relationship between days after 50% heading and the relative yield observed in the shooting date experiment. For the data collected at M'bé, Cote d'Ivoire.

$$y = a + bx + b(c_1 - x) \cdot I_{(x-c_1)} \quad (2)$$

and for the data collected at Marovoay, Madagascar;

$$y = a + b(x - c_1) \cdot I_{(x-c_1)} + b(c_2 - x) \cdot I_{(x-c_2)} \quad (3)$$

were used, respectively. The parameters  $a$  and  $b$  are constant,  $y$  is the ratio between the observed and the final yield, and  $x$  is the date after 50% heading. The parameters  $c_1$  and  $c_2$  are the breaking points of the segments, and Eq. (3) represents the 3 segmented regression. Function 'I' is the step function, which is defined as follows:

$$I_{(x)} = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad (4)$$

For the dataset in Madagascar for the shooting date experiment, the six estimations from 1200 to 1250 hrs were averaged and defined as an estimation for a plot. The estimations at seven harvested plots were then averaged, and the standard error was calculated. All analyses in the present study were conducted using Microsoft Excel (Microsoft, Redmond, WA, USA), Neural Network Console software (Sony Network Communications Inc., Japan), and Python language version 3.7 (<http://www.python.org>) with Pytorch framework version 1.7 (<https://pytorch.org/>). The code to run the developed CNN model is available at [https://github.com/r1wtn/rice\\_yield\\_CNN.git](https://github.com/r1wtn/rice_yield_CNN.git).

### **Data Availability**

The data that support the findings of this study are available from the authors on reasonable request.

## **Declarations**

### **Acknowledgements**

We are grateful for the financial support to this study by the European Union and International Fund for Agricultural Development (IFAD) under the project “Sustainable and Diversified Rice-based Farming Systems [DCIFOOD/2015/360-968]” under the program “Putting Research into Use for Nutrition, Sustainable Agriculture and Resilience (PRUNSAR).” and the CGIAR Research Program (CRP) on rice agri-food systems (to K. S.), JSPS KAKENHI Grant Number: JP19H02939 and 20H02968, the Ministry of Agriculture, Forestry and Fisheries of Japan [Smart-breeding system for Innovative Agriculture] and Cabinet Office, Government of Japan [PRISM] (BAC1003) (to Y. Tanaka), JICA/JST SATREPS, Japan Grant No. JPMJSA1608 (to Y. Tsujimoto).

### **Author contributions**

Y. T. and K. S. designed and performed research and visualisation, and wrote the paper; Y.T. and K.S. analysed the data and results; all authors collected data and contributed to data collection and editing.

### **Competing financial interests**

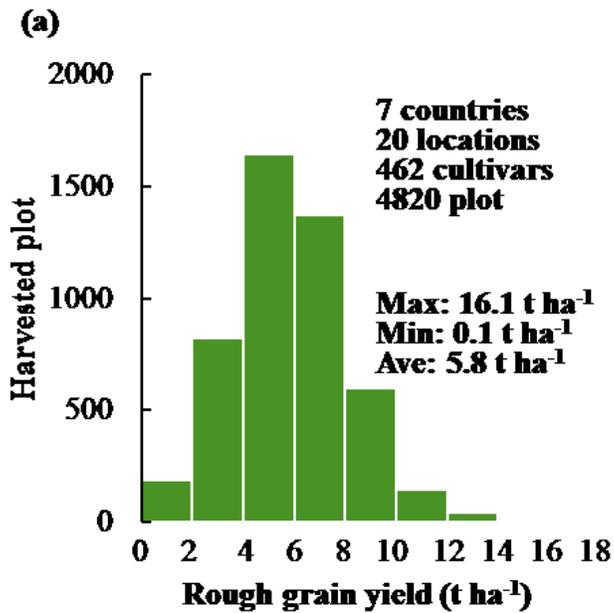
The authors declare no competing financial interests

## References

1. Godfray HC, et al. Food security: The challenge of feeding 9 billion people. *Science* 327(5967), 812–818. (2010).
2. Ramankutty, N., Foley, J. A., Norman, J. and McSweeney, K. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Glob. Ecol. Biogeogr.* 11(5), 377–92. (2002).
3. Fischer, R. A., Byerlee, D. and Edmeades, G. O. Crop yields and global food security: will yield increase continue to feed the world? *ACIAR Monograph No. 158*. Australian Centre for International Agricultural Research: Canberra. xxii + 634 pp. (2014).
4. Saito, K., Six, J., Komatsu, S., Snapp, S., Rosenstock, T., Arouna, A., Cole, S., Taulya, G., S. Vanlauwe, B. Agronomic gain: definition, approach and applications. *Field Crops Res.* 270, 108193. (2021).
5. Bruke, M., Lobell, D.B. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci. U.S.A.* 114, 2189–2194. (2017).
6. Lobel D.B. et al. Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis. *Amer. J. Agr. Econ.* 102, 202–219. (2019).
7. Lecun, Y., Bengio, Y., Hinton, G. Deep learning. *Nature* 521, 436–444. (2015).
8. Nature. The scientific events that shaped the decade *Nature* 576, 337-338. (2019).
9. Popel, M. et al. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature commun.* 11, 4381. (2020).
10. Senior, A.W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710. (2020).
11. Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* 550, 354–359. (2017).
12. Liang, W., Zhang, H., Zhang, G., Cao H. Rice blast disease recognition using a deep convolutional neural network. *Sci. Rep.* 9, 2869. (2019).
13. Sharma, P., Berwal, Y.P.S., Ghai, W. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation *Information Processing in Agriculture.* 7, 566–574. (2020).
14. Rustia, D.J.A. et al. Automatic greenhouse insect pest detection and recognition based on a cascaded deep learning classification method. *J. Applied Entomology.* 145, 206–222. (2021).
15. Ghosal, S. et al. An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4613-4618. (2018)
16. Ma, J. et al. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *European J. Agronomy* 103, 117–129. (2019).
17. Castro, W. et al. Deep learning applied to phenotyping of biomass in forages with UAV-based RGB imagery. *Sensors* 20, 4802. (2020).

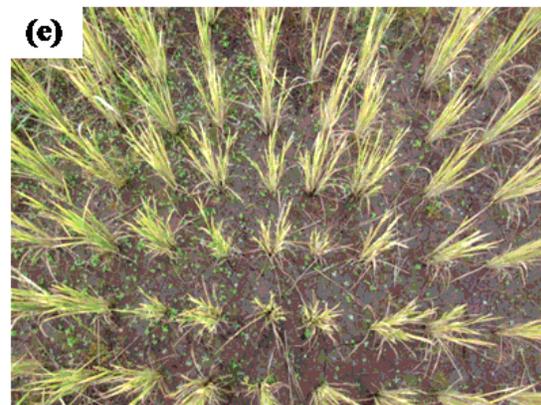
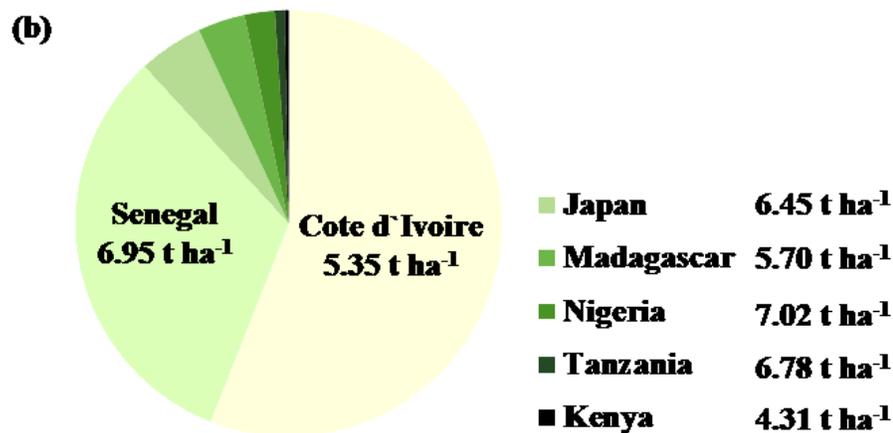
18. Jin, X., Li, Z., Feng, H., Ren, Z., Li, S. Deep neural network algorithm for estimating maize biomass based on simulated Sentinel 2A vegetation indices and leaf area index. *The Crop J.* 8, 87–97. (2020).
19. Gen, L., Che, T., Ma, M., Tan, J., Wang, H. Corn biomass estimation by integrating remote sensing and long term observation data based on machine learning techniques. *Remote Sensing* 13, 2352. (2021).
20. Apolo-Apolo, O.E., Pérez-Ruiz, M., Martínez-Guanter, J., Egea, G. A. Mixed Data-Based Deep Neural Network to Estimate Leaf Area Index in Wheat Breeding Trials. *Agronomy* 10, 175. (2020).
21. Toda, Y. et al. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Commun. Biology* 3, 1–12. (2020).
22. Xiong, X. et al. Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* 13, 104. (2017).
23. David, E. et al. Global Wheat Head Detection (GWHD) Dataset: A Large and Diverse Dataset of High-Resolution RGB-Labelled Images to Develop and Benchmark Wheat Head Detection Methods. *Plant Phenomics* 3521852. (2020).
24. GRiSP (Global Rice Science Partnership). *Rice almanac, 4th edition*. Los Baños (Philippines): International Rice Research Institute. 283 p. (2013).
25. Zeiler M.D., Fergus R. Visualizing and Understanding Convolutional Networks. *arXiv1311.2901v3* (2013).
26. Lobell D.B. The use of satellite data for crop yield gap analysis. *Field Crops Res.* 143, 56–64. (2013).
27. Setiyono T.D. et al. Rice yield estimation using synthetic aperture radar (SAR) and the ORYZA crop growth model: development and application of the system in South and South-east Asian countries. *International Journal of Remote Sensing* 40, 8093–8124, (2019).
28. Jain, M. et al. Using satellite data to identify the causes of and potential solutions for yield gaps in India's Wheat Belt. *Environ. Res. Lett.* 12, 094011. (2017).
29. Lobell, D.V. et al. Sight for Sorghums: Comparisons of Satellite- and Ground-Based Sorghum Yield Estimates in Mali. *Remote Sensing*. 12, 100. (2020).
30. Zhou, X. et al. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *J. Photogrammetry and Remote sensing* 130, 246–255. (2017).
31. Wang et al. Application of UAS in crop biomass monitoring: A review. *Front. Plant Sci.* 12, 616689.
32. Li, R., Li, M., Ashraf, U., Liu, S., Zhang, J. Exploring the relationships between yield and yield-related traits for rice varieties released in China from 1978 to 2017. *Front. Plant Sci.* 10, 543. (2019).
33. He, K., Zhang, X., Ren, S., Sun J. Deep residual learning for image recognition. *arXiv.1512.03385v1* (2015).
34. FAO. *Guidelines on Planning Rice Production Survey*. Rome. 168 pp. (2019).

## Figures



**(c)**

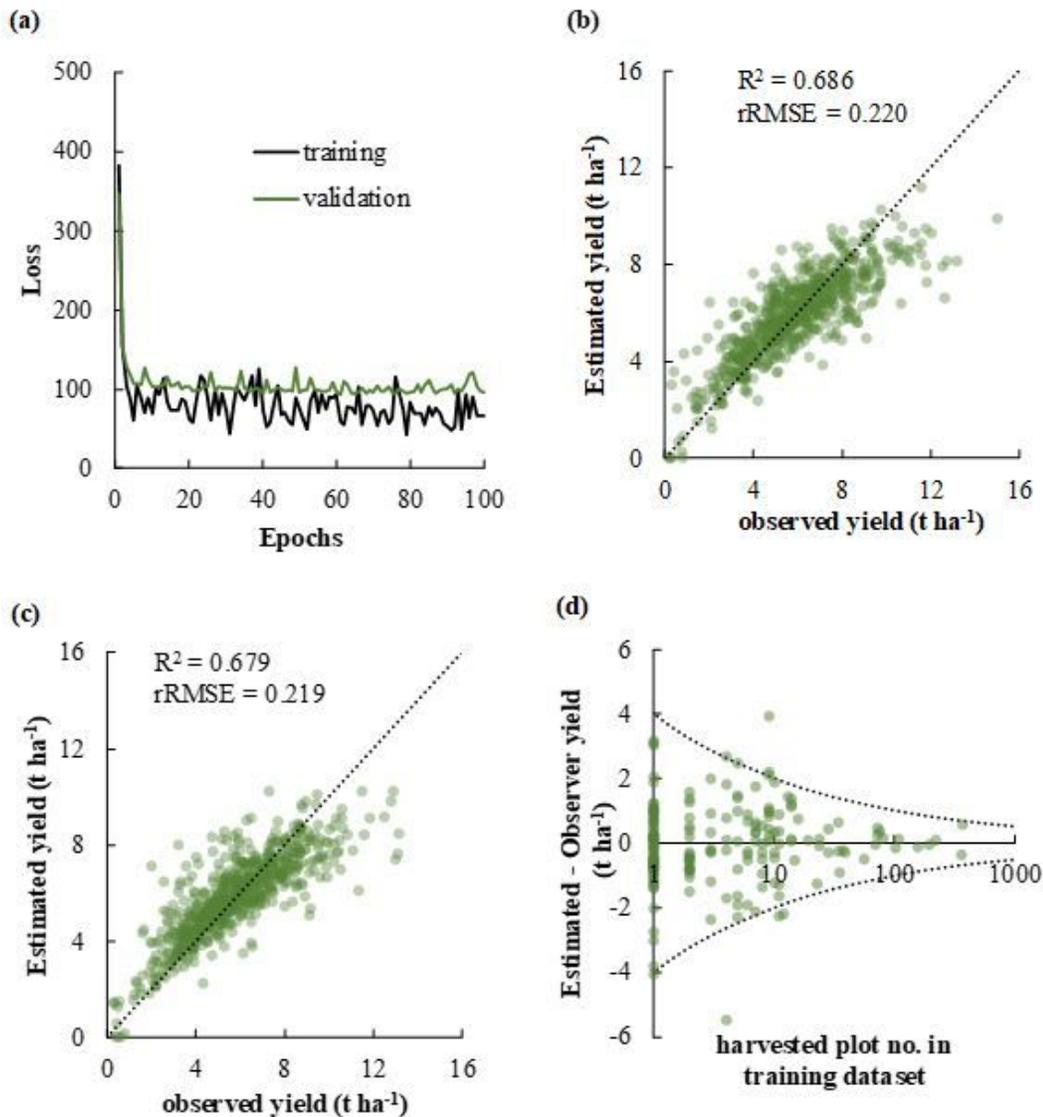
	Category	Harvested plot	Image no.
development and evaluation	training	3381	17764
	validation	661	697
	test	654	689
	panicle removal	10*	50
robustness	angle	25	1000
	time of day	1	114
	shooting date	41	1518
prediction	prediction	47	235
<b>Total</b>		<b>4820</b>	<b>22067</b>



**Figure 1**

The global database of the rice canopy image and corresponding rough grain yield (a) Bar plot depicting the frequency distribution of observed rough grain yield in the database collected at 7 countries. (b) Pie-chart of the dataset composition of 7 countries with the average rough grain yield in each country. (c) Tabular representation of the dataset classified into 3 major parts; development and evaluation, robustness, and prediction. For the sub-categories, the training and validation dataset is directly used to

develop the model. The test and panicle removal datasets were used to evaluate the accuracy and characteristics of the model. For details of angle, time of day, and shooting date, please refer the main text. The prediction dataset consists of the data at Tanzania and Japan (Tokyo University of Agriculture and Technology), which are not included to any other categories. \*: 10 different yield data were generated for single plot of cv. Koshihikari at Kyoto, Japan by the sequential panicle removal. (d-e) Images of the rice canopy showing the highest (d) and the lowest (e) rough grain yield, which is collected in Senegal and Madagascar, respectively.



**Figure 2. Development of the deep learning-based model to estimate the rough grain yield of rice**

(a) Graph illustrating the learning curve for the training and validation dataset. The minimum loss for the validation dataset was recorded at epoch = 61, and the model was used for all the analyses in the present study. (b-c) Scatter plots depicting the estimation of the rough grain yield in the validation (b) and the test (c) dataset. The dotted line represents the 1:1 relationship. rRMSE represents the relative root mean square error. (d) Scatter plot showing the difference between estimated and observed yield of each cultivar in test dataset plotted against the number of harvested plots in training dataset.

The dotted line represents  $y = \pm 4 \cdot \left(\frac{1}{2}\right)^{\log_{10} x}$

Figure 2

See image above for figure legend.

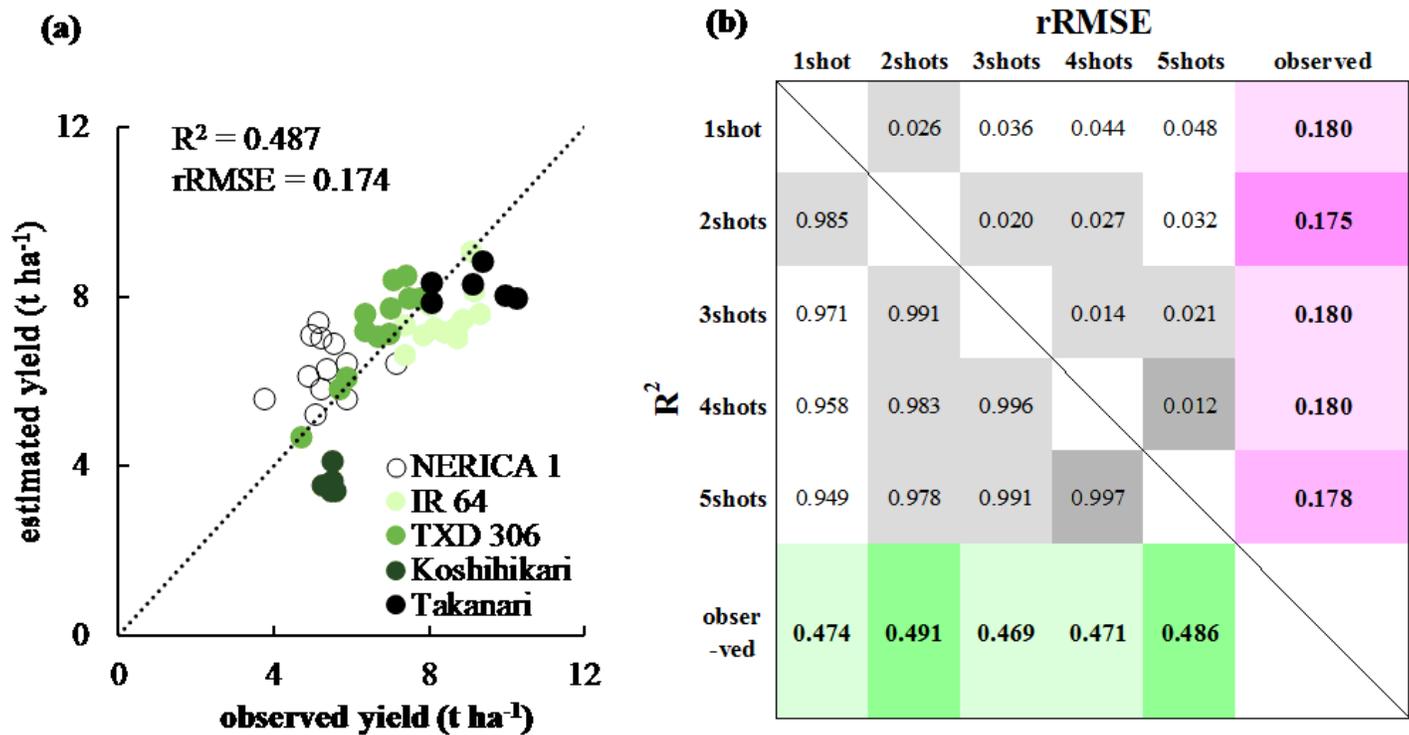


Figure 3

The prediction of the rough grain yield of datasets independent from the model establishment. (a) Scatter plot of the relationship between observed and estimated yield of the prediction datasets. NERICA1, IR 64 and TXD 306 were harvested in Tanzania. Koshihikari and Takanari were harvested in Tokyo, Japan. Both of the locations were not included in any other model analyses. (b) The matrix of rRMSE and R2 with the observed and estimated yield for prediction dataset with different number of replicated shots per harvested plot.

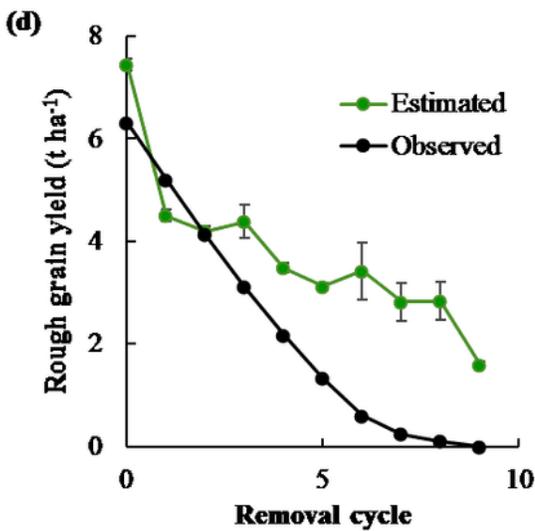
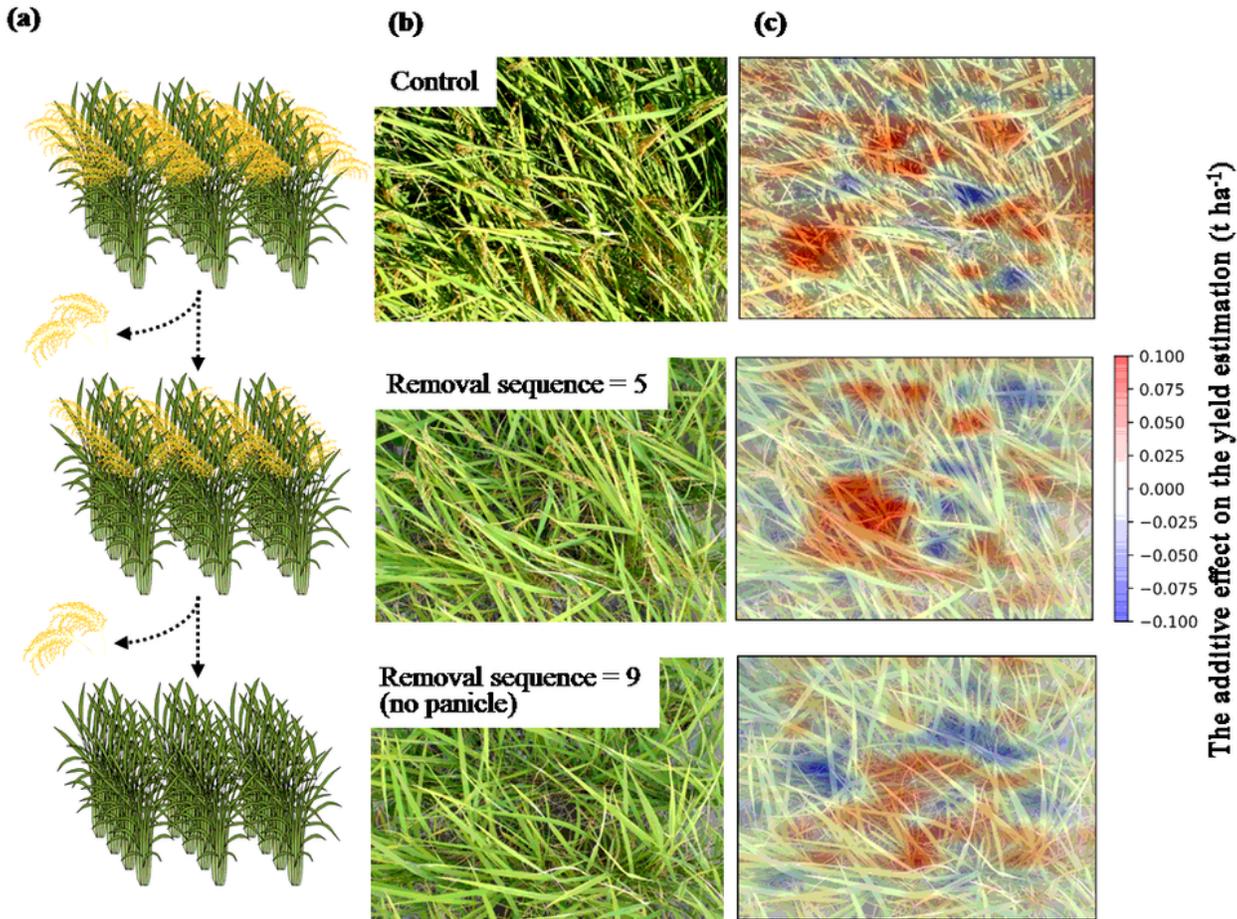
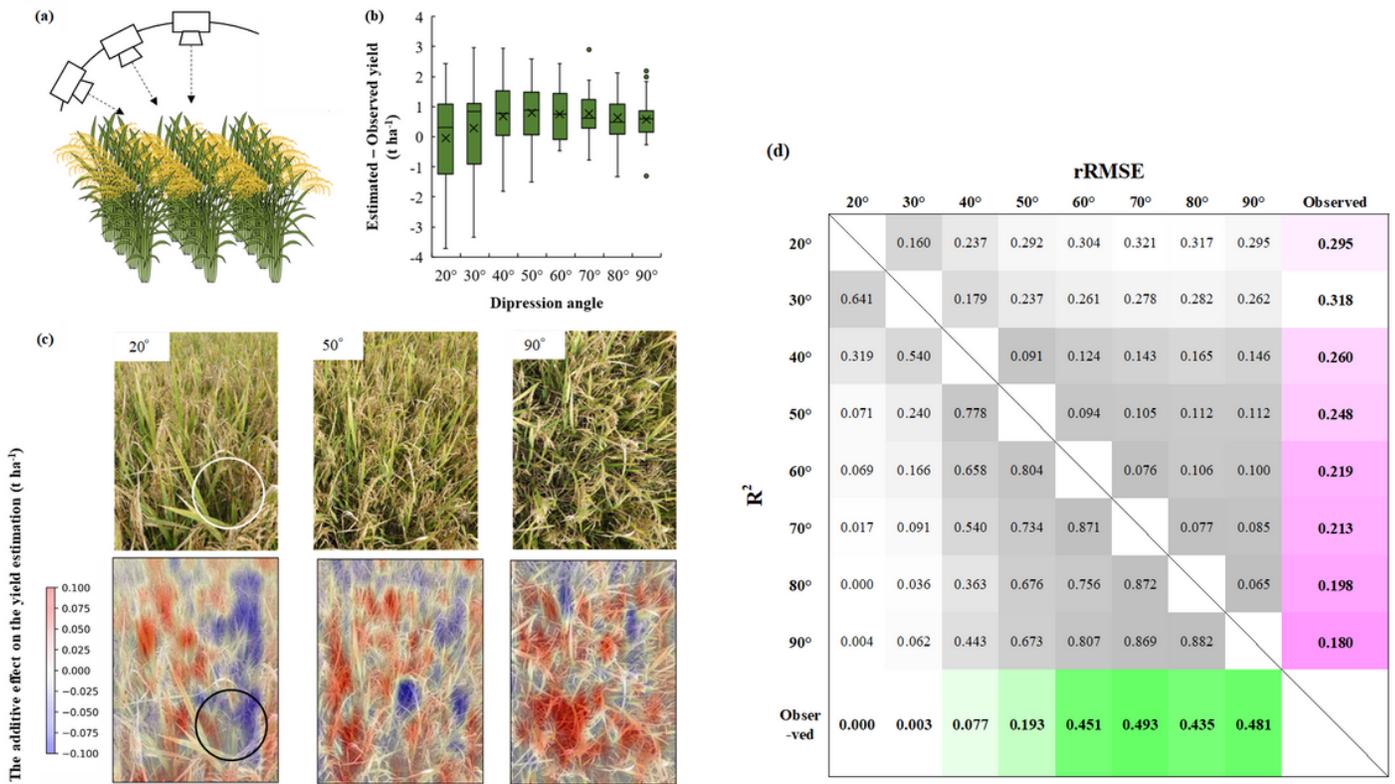


Figure 4

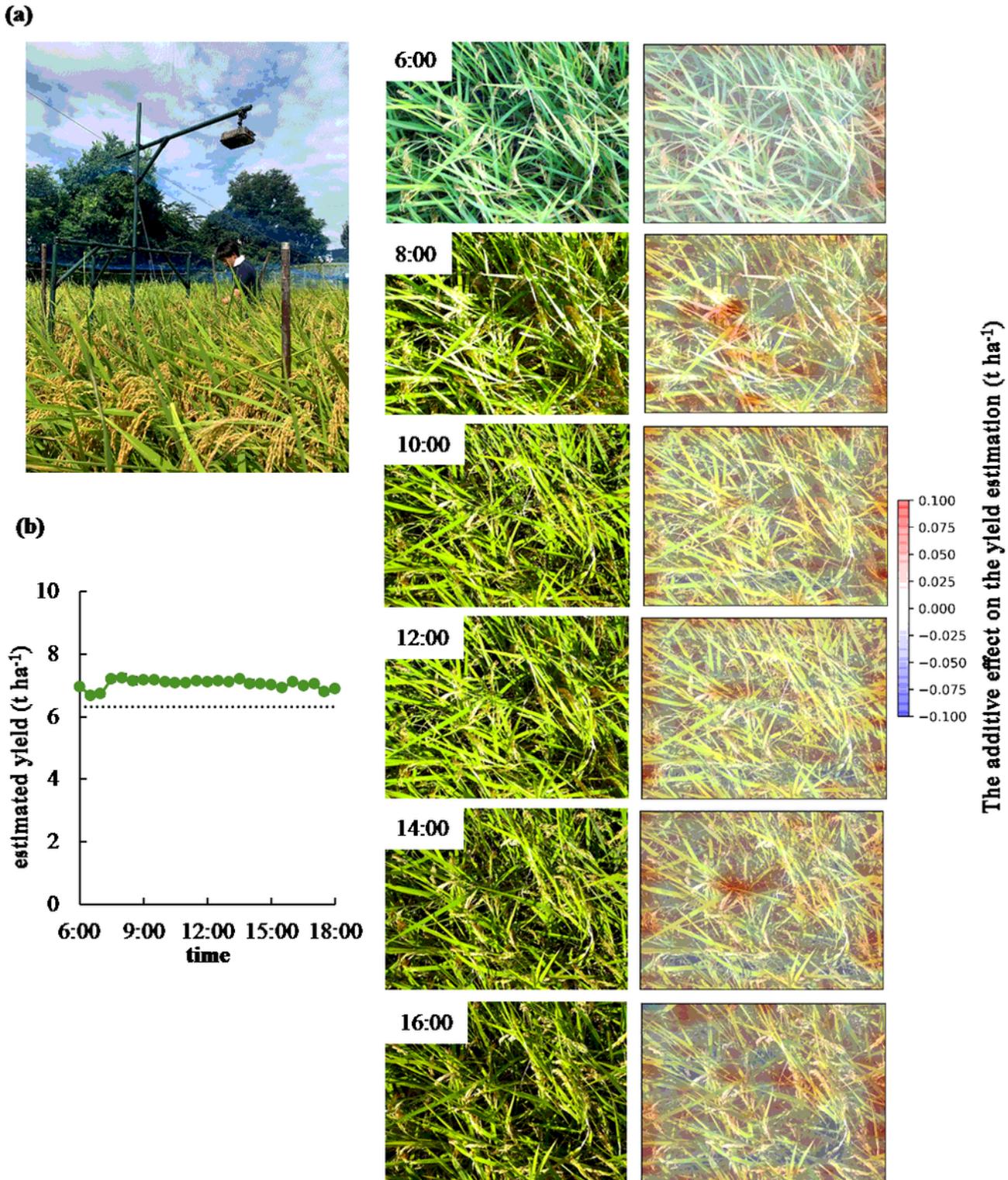
Panicle removal experiment to visualize the quantitative contribution of the panicles to the yield estimation. (a) Schematic illustration of the panicle removal. The image was taken at the fixed camera above the rice canopy (cv. Koshihikari). Two panicles per plant were removed and the image was taken for 5 times with the random shift. This cycle was repeated until all the panicles were removed from the canopy. (b-c) The canopy image and corresponding heatmap of the additive effect on the yield

estimation before panicle removal, removal cycle = 5 and 9 (no panicle). For the technical detail of the visualization, see Figure S5. (d) Line graph of the observed and the estimated yield plotted against the removal cycle. The bars represent the standard error (n = 5).



**Figure 5**

The effect of the depression angle on the yield estimation. (a) Schematic illustration of the protocol to take the image with various depression angles. The digital camera was sided on the rail above the rice canopy to change the depression angle with the same rice canopy. The image at each angle was taken for 5 times with the random shift. (b) Boxplot of the relative error to the observed yield against the depression angle of the image. (n = 25). (c) Heatmap of the additive effect on the yield estimation with original images taken with the depression angles of 20, 50, and 90 degrees (default protocol). For the technical detail of visualization, refer to Figure S5. The white and black circles represent the regions of the inner structure of the canopy with the negative effects on yield estimation. (d) The matrix of rRMSE and R<sup>2</sup> with the observed and estimated yield for images taken from various depression angles. The depression angle of 90° is identical to the default protocol.



**Figure 6**

The effect of time of day on the yield estimation. (a) Image showing the digital camera fixed above the rice canopy (cv. Koshihikari), and the image was taken every 30 min from 0600 to 1800 hrs. The camera was fixed at 5 days before harvesting and the image was collected until the harvest day. (b) Plot of the estimated rough grain yield against the time of day. The data of 6 days were averaged and the standard error is represented by bars ( $n = 5$ ). The dotted line represents the observed rough grain yield of 6.31 (t ha<sup>-1</sup>)

1) (c-d) Images recorded at 0600, 0800, 1000, 1200, 1400, and 1600 hrs on Aug. 29, 2020 (c) and the corresponding heat maps of the additive effect on the yield estimation (d).

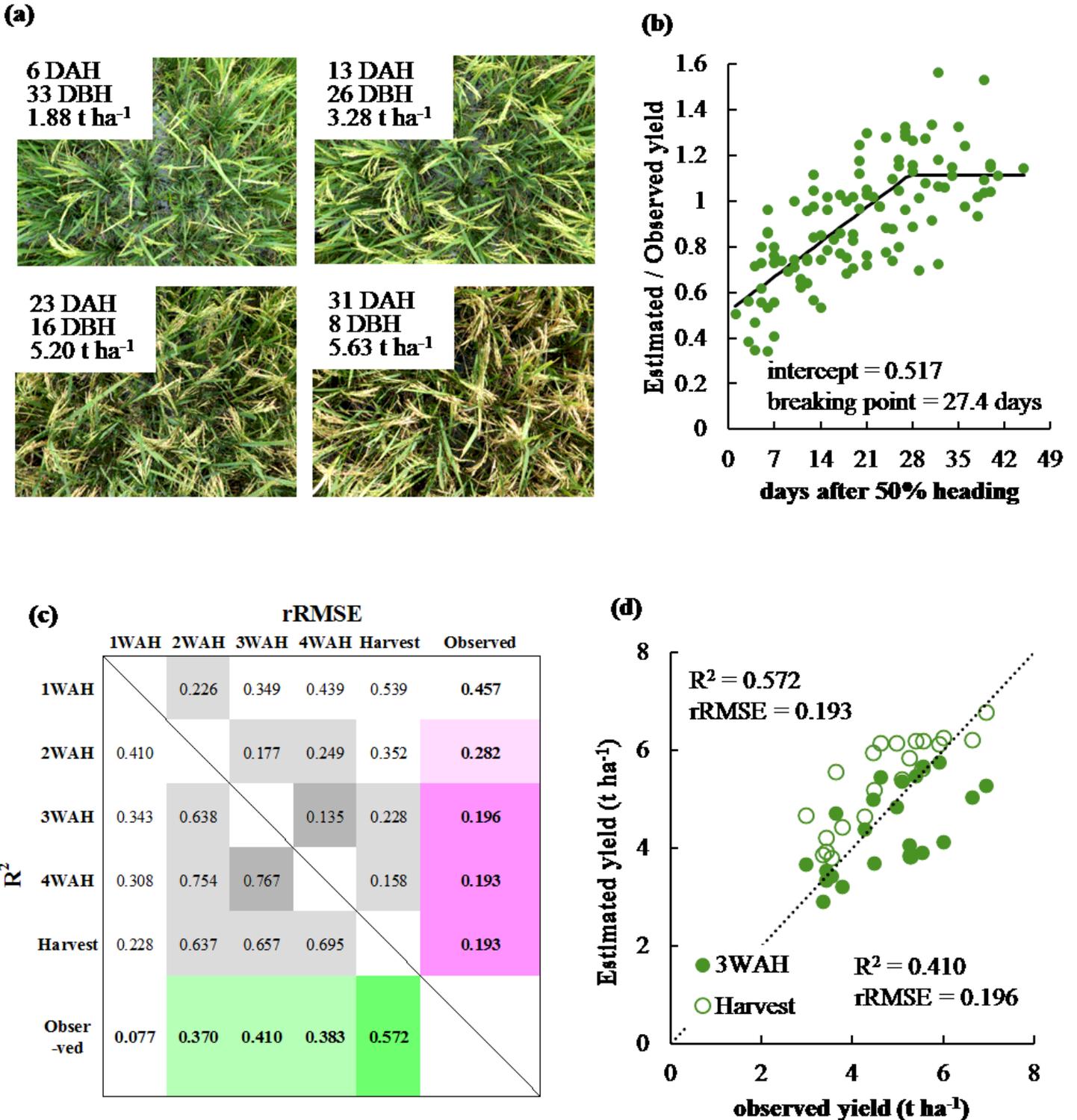


Figure 7

The applicability of the model to the images taken during grain filling stage. (a) Examples of the image taken at approximately 1 to 4 weeks after 50% heading (cv. IRRI 154). (b) Scatter plot of the estimated yield relative to the final yield plotted against days after 50% heading. The data consists of images of 22

cultivars taken at 1 to 4 weeks after 50% heading. The solid line represents the segmented linear regression. (c) The matrix of rRMSE and R<sup>2</sup> with the observed and estimated yield for images taken at approximately 1 to 4 weeks after 50% heading and at harvest. (d) Scatter plot depicting the relationship between observed and estimated yield at 3 weeks after 50% heading (3WAH) and at harvest. The dotted line represents the 1:1 relationship.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1.xlsx](#)
- [SupplementaryFigures.pdf](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS3.xlsx](#)