Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

RiGaD: An aerial dataset of rice seedlings for assessing germination rates and density

Trong Hieu Luu[a], Hoang-Long Cao[a,c], Quang Hieu Ngo[a,*], Thanh Tam Nguyen[b], Ilias El Makrini[c], Bram Vanderborght[d]

[a] College of Engineering, Can Tho University, Can Tho 910900, Viet Nam
[b] Mekong Development Research Institute, Can Tho University, Can Tho 910900, Viet Nam
[c] Brubotics, Vrije Universiteit Brussel and Flanders Make, Brussels 1050, Belgium
[d] Brubotics, Vrije Universiteit Brussel and imec, Brussels 1050, Belgium

A R T I C L E I N F O

Article history:
Received 13 April 2024
Revised 19 June 2024
Accepted 31 October 2024
Available online 6 November 2024

Dataset link: RiGaD: An aerial dataset of rice seedlings for assessing germination rates and density Creators (Original data)

Keywords:
Rice
Seedlings
Sowing
UAV
Image processing
Dataset
Machine learning

A B S T R A C T

The popularity of Unmanned Aerial Vehicles (UAVs) in agriculture makes data collection more affordable, facilitating the development of solutions to improve agricultural quality. We present a dataset of rice seedlings extracted from aerial images captured by a UAV under various environmental conditions. We focus on rice seedlings cultivated by the sowing method during their early growth stages because these stages are important to the establishment and survival as well as foundation for lifelong growth. We employed an adaptive thresholding method to isolate rice seedlings from the aerial images. We subsequently classified them into three categories based on their germination conditions: single rice seedings, clustered rice seed plants, and undefined objects. We obtained a total of 5364 labeled images of rice seedlings through data augmentation. This dataset serves as a resource for assessing germination rates and density using machine learning methods. The results derived from these assessments help farmers understand seedling growth and enable them to monitor the health and vigor of rice seedling during early growth stages.

* Corresponding author.
E-mail address: nqhieu@ctu.edu.vn (Q.H. Ngo).
Social media: @hoanglongcao?s=21 (H.-L. Cao)

https://doi.org/10.1016/j.dib.2024.111118

Specifications Table

Subject	Agricultural Sciences
Specific subject area	Machine learning, deep learning, image detection and classification, agriculture engineering, robotics
Type of data	Image, Raw, Labelled, Augmented
Data collection	We used a DJI-FC6310S camera, which provides high-resolution images (4864 × 3648 pixels), mounted on a Phantom Pro 4 UAV to collect aerial images. The data collection process targeted the rice variety IR4625, 17 days after sowing, on June 6th, 2023, during the summer-fall crop season. In total, aerial images were collected, covering various environmental conditions. The rice seedling images were extracted from these aerial images and classified based on their germination conditions: single rice seed plants, clustered rice seed plants, and undefined objects. The dataset comprises a total of 5,364 images (224 × 224 pixels), with a combined size of 35 MB and 10 raw photos taken from UAV
Data source location	Institution: College of Engineering, Can Tho University Sampled paddy field: Thu Thua, Long An, Vietnam Latitude, longitude for collected data samples: 10°37′13.3"N 106°20′06.1"E Altitude: 8 meters
Data accessibility	Repository name: RiGaD: An aerial dataset of rice seedlings for assessing germination rates and density Data identification number: 10.5281/zenodo.11658969 Direct URL to data: https://zenodo.org/records/11658969 Instructions for accessing these data: users can download directly via URL, extract and use it.
Related research article	N/A

1. Value of the Data

- The dataset focuses on rice seedlings cultivated using the sowing method during their early growth stages, distinguishing it from other datasets that may focus on transplanting or later growth stages.
- The dataset was generated from aerial images taken on a UAV at an altitude of 8 meters, which include rice seedlings at various environmental conditions such as backlit, normal light, dark background, and underwater.
- Researchers can use this dataset for assessing germination rate and density of rice seedlings using machine learning and deep learning methods.

2. Background

Understanding rice seedling growth at early stages, such as germination rate and density, enables farmers to effectively monitor the health and vigor of rice plants. These assessments traditionally rely on manual sampling and statistics, where farmers count the total number of rice plant tillers in a small sample area. This process is time-consuming and labor-intensive. **Automatic assessment** has emerged as a recent solution **by applying machine learning trained on rice seedling datasets.**

The currently available datasets vary on the growth stages. Rice plant growth can be divided into three phases of development including vegetative, reproductive and ripening (Fig. 1A). At the vegetative phase, some datasets focused on transplanting plants [1–4]. At the reproductive and ripening stages, some dataset focused on plants health evaluation from tillering to ready harvesting [5], rice lodging estimation [6], and rice grain yield [7]. Other datasets used time

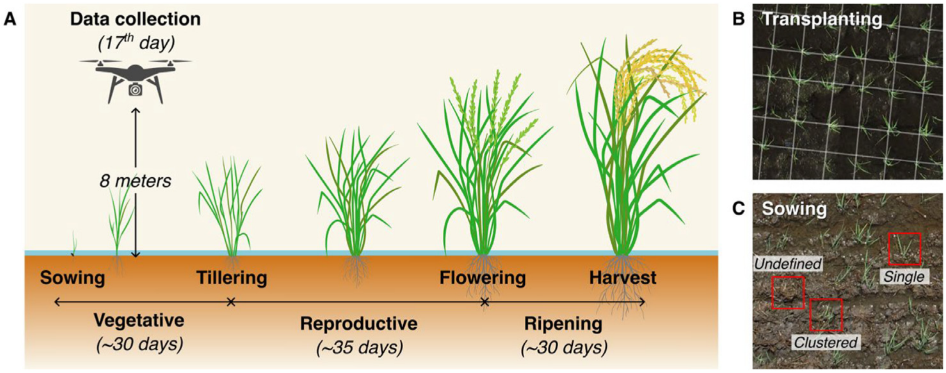

Fig. 1. Data collection (A) was performed at the early growth stages of rice cultivated by the sowing method, using UAV technology. Compared to the transplanting method (B), the sowing method (C) led to a more heterogeneous plant distribution, particularly marked by an increased incidence of clustered rice seedlings.

series to monitor rice health and density from transplanting to tillering [8], water management [9], phenological detection [10–12], and monitoring all development stages [13].

Most datasets were generated from **transplanted plants**, wherein rice seedlings undergo the germination process in nursery beds and are later planted into rows (Fig. 1B). However, sowing (Fig. 1C) is also a popular method, where rice seeds are planted directly onto the field surface, typically with three to five seeds per hole. This practice often results in a more heterogeneous plant distribution, characterized by an increased incidence of clustered rice seedlings. There is a need for **more datasets generated from rice seedlings cultivated by the sowing method**.

In the work, **we generated a dataset of cultivated by the sowing method during their early growth stages**. First, we used an Unmanned Aerial Vehicle (UAV) to capture aerial images under various environmental conditions. Second, we applied an adaptive thresholding method to isolate rice seedlings and classified them into three categories based on their germination conditions: single rice seed plants, clustered rice seed plants, and undefined objects. Finally, we used data augmentation to increase the randomness of the dataset. This dataset provides a resource for assessing germination rates and density using machine learning techniques.

3. Data Description

Data was collected during the summer-fall crop season in Thu Thua, Long An province, Vietnam. The sampled paddy field is located at 10°37′13.3"N 106°20′06.1"E (Fig. 2). This season is

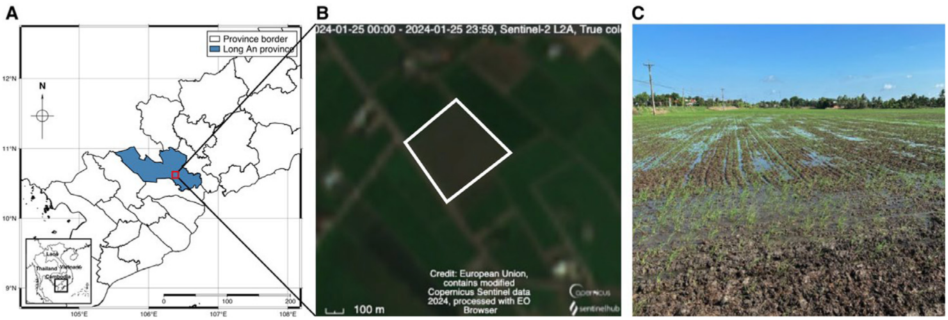

Fig. 2. Location of the sampled paddy field in Long An province, Vietnam.

characterized by hot and humid weather nationwide. Flooding and waterlogging from heavy rainfall and typhoons can lead to crop damage. The dataset contains 5,364 images (224 × 224 pixels) of rice seedlings with a combined size of 35 MB and 10 raw photos from UAV. These images were manually labeled and organized in three folders: single rice seed plants (1367 images), clustered rice seed plants (1437 images), and undefined objects (2560 images) (Fig. 3).

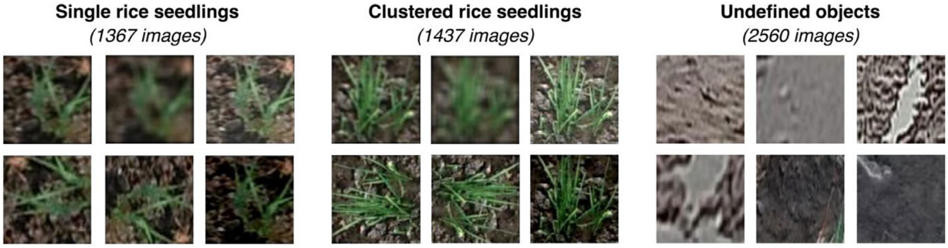

Fig. 3. Example images from the dataset showcasing three categories: single rice seedlings, clustered rice seedlings, and undefined objects.

4. Experimental Design, Materials and Methods

4.1. Aerial Data Acquisition

We used a DJI-FC6310S camera mounted on a Phantom Pro 4 UAV to capture high-resolution images (4864 × 3648 pixels) from a height of 8 meters. The rice variety was IR4625, 17 days after sowing, on June 6th, 2023, during the summer-fall crop season. In total, 211 aerial images were collected, covering various environmental conditions: normal light, backlit, dark background, and underwater (Fig. 4). Detailed of environmental condition is described as follows.

17 photos have dark background and under water condition,
40 photos have normal background and under water condition,
37 photos have normal background and backlit,
8 photos are with inverted view of the seedling in normal condition,
13 photos are with inverted view of the seedling in backlit and water condition,
29 photos are with normal conditions,
32 photos are with underwater conditions and backlit,
35 photos are with normal, underwater and backlit conditions.

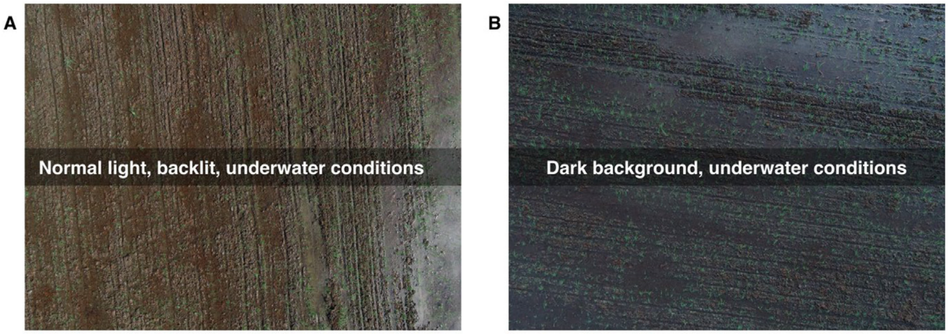

Fig. 4. Aerial data acquisition under various light conditions: normal light, backlight, dark background, and underwater.

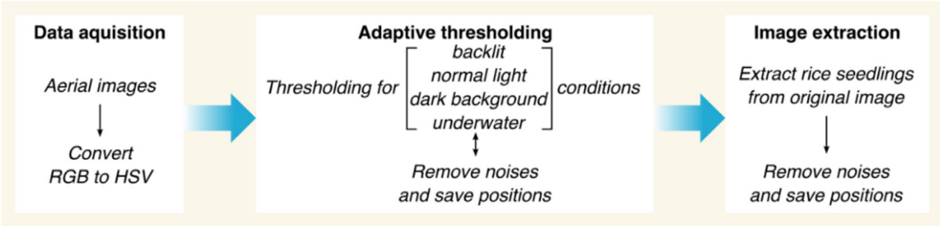

Fig. 5. Rice seedlings extraction from area images using an adaptive thresholding method for different light conditions: backlit, normal light, dark background, and underwater.

4.2. Rice Seedlings Extraction from Aerial Images

At first, We measured the specification of the camera and we get the value that the ground distance sample (GSD) at 8 meters height is at 0.2 cm/px. Next, an adaptive image thresholding method was applied to extract rice seedling from aerial images (Fig. 5). Specifically, each aerial image was first converted from RGB to HSV, followed by implementation of adaptive threshold parameters in four main cases: backlit, normal light, dark background, and underwater (Fig. 6). These parameters were selected through iterative trial-and-error testing. At each thresholding step, we removed uncertain noises by excluding objects smaller than 135 $\times$ 135 pixels. In detailed, we proposed 12 parameters of adaptive thresholding as follows.

- Parameter 1: minimum (32, 28, 0), maximum (57, 255, 255)
- Parameter 2: minimum (30, 245, 0), maximum (76, 186, 186)
- Parameter 3: minimum (31, 0, 0), maximum (64, 100, 180)
- Parameter 4: minimum (21, 48, 48), maximum (106, 151, 151)
- Parameter 5: minimum (45, 30, 30), maximum (89, 115, 190)
- Parameter 6: minimum (47, 0, 82), maximum (85, 255, 255)
- Parameter 7: minimum (58, 0, 72), maximum (95, 255, 255)
- Parameter 8: minimum (47, 0, 82), maximum (85, 255, 255)
- Parameter 9: minimum (47, 0, 82), maximum (85, 255, 255)
- Parameter 10: minimum (39, 13, 80), maximum (90, 255, 255)
- Parameter 11: minimum (0, 0, 104), maximum (110, 157, 255)
- Parameter 12: minimum (35, 0, 0), maximum (179, 255, 255)

At 135 $\times$ 135 pixels, the size of the object is about 27 cm. At this size, the object should be the seedlings. At the size under 10 $\times$ 10 pixels, it should be noises from light, sun glitter from environment, so we remove these sizes. Once the thresholding positions were marked, we extracted rice seedlings from the original aerial image for the image labelling process. Among the noise, we classified those larger than 10 $\times$ 10 pixels as undefined objects such as straws, sand, soils, rotten roots, trash. At the size of higher than 10 $\times$ 10 pixels, it can be undefined objects (as the photos), so we have to keep these objects and choose again manually.

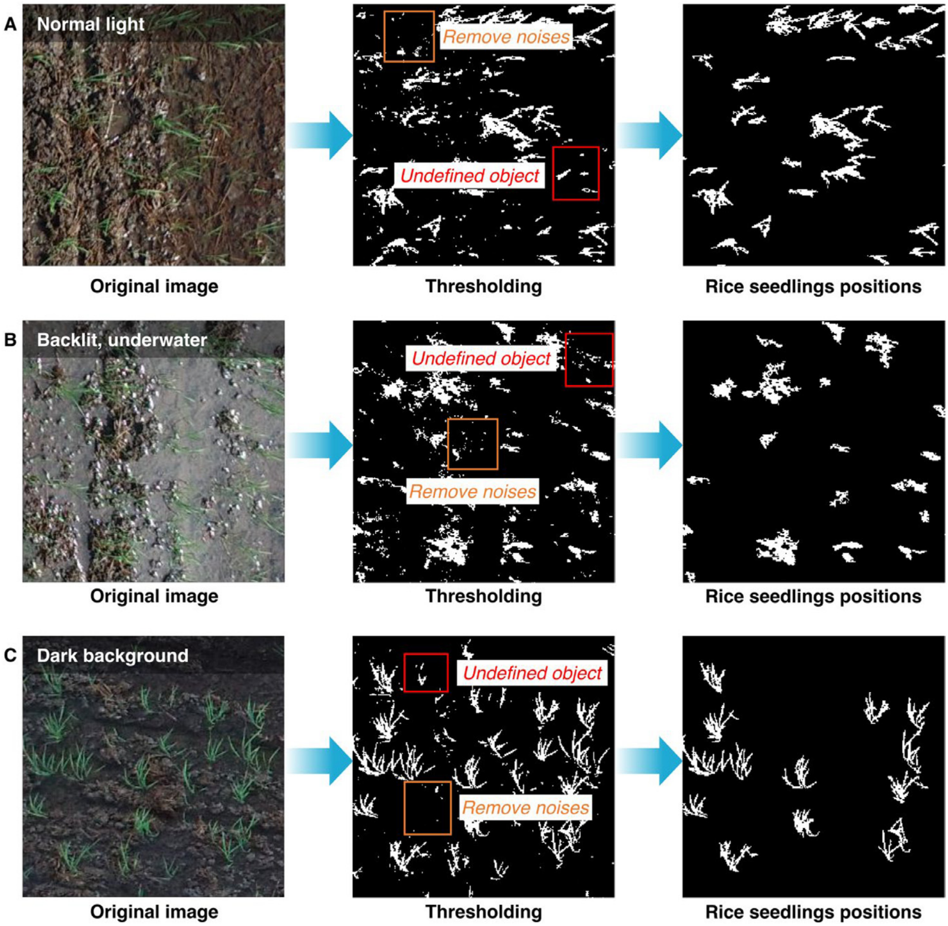

Fig. 6. Example of rice seeding detection and extraction in different light conditions. (A) Normal light. (B) Backlit, underwater. (C). Dark background.

4.3. Image Labelling and Data Augmentation

After extracting rice seedlings from original aerial images, we classified them (manual labelling) into three categories: single rice seedings, clustered rice seedlings, and undefined objects. We applied data augmentation on single and clustered rice seedlings images to enhance the randomness of rice seedling conditions, using image enhancement methods such as rotation, blurring, and brightness adjustments (Fig. 7). Specifically, we rotated images by 90 and -90 degrees to account for vertical camera orientations during flight, blurred images using Gaussian filters of sizes [3,3] and [5,5], and adjusted lighting by altering gray values. No data augmentation was applied to undefined objects due to the large number of images. All images were standardized to a resolution of 224 $\times$ 224 pixels.

Fig. 7. Data augmentation was applied on original images of single rice seedlings (A) and clustered rice seedling (B) to improve the randomness of the dataset.

Limitations

The aerial images were captured during a single crop season. Expanding the dataset in various weather conditions and in different crop seasons would enhance its overall diversity.

Ethics Statement

The proposed dataset does not involve any human subjects, animal experiments, or data collected from social media or other sources.

Data Availability

RiGaD: An aerial dataset of rice seedlings for assessing germination rates and density Creators (Original data) (Zenodo).

CRediT Author Statement

Trong Hieu Luu: Conceptualization, Methodology, Resources, Data curation, Writing – original draft, Validation; **Hoang-Long Cao:** Methodology, Writing – original draft, Visualization, Supervision; **Quang Hieu Ngo:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration; **Thanh Tam Nguyen:** Data curation, Validation, Writing – review & editing; **Ilias El Makrini:** Writing – review & editing, Supervision, Funding acquisition; **Bram Vanderborght:** Writing – review & editing, Supervision, Funding acquisition.

Acknowledgements

The authors acknowledge that this work was supported by the Brussels Institute of Advanced Studies (Grant: BrIAS2024) and by a scientific stay grant from the FWO (Grant number: V501724N).

This study is also funded by the Ministry of Education and Trainning Project with the code number B2023.TCT.08.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Yang, H. Tseng, Y. Hsu, C. Yang, M. Lai, D. Wu, A UAV open dataset of rice paddies for deep learning practice, Remote Sens. 13 (2021) 1358, doi:10.3390/rs13071358.
[2] H.H. Tseng, M.D. Yang, R. Saminathan, Y.C. Hsu, C.Y. Yang, D.H. Wu, Rice seedling detection in UAV images using transfer learning and machine learning, Remote Sens. 14 (2022) 2837, doi:10.3390/rs14122837.
[3] J. Wu, G. Yang, X. Yang, B. Xu, L. Han, Y. Zhu, Automatic counting of *in situ* rice seedlings from UAV images based on a deep fully convolutional neural network, Remote Sens. 11 (2019) 691, doi:10.3390/rs11060691.
[4] X. Ma, X. Deng, L. Qi, Y. Jiang, H. Li, Y. Wang, X. Xing, Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields, PLoS ONE 14 (2019), doi:10.1371/journal.pone.0215676.
[5] X. Bai, P. Liu, Z. Cao, H. Lu, H. Xiong, A. Yang, Z. Cai, J. Wang, J. Yao, Rice plant counting, locating, and sizing method based on high-throughput UAV RGB images, Plant Phenomics 5 (2023) Article 0020, doi:10.34133/plantphenomics.
[6] M.D. Yang, J.G. Boubin, H.P. Tsai, H.H. Tseng, Y.C. Hsu, C.C. Stewart, Adaptive autonomous UAV scouting for rice lodging assessment using edge computing with deep learning EDANet, Comput. Electron. Agric. 179 (2020) 105817.
[7] Q. Yang, L. Shi, J. Han, Y. Zha, P. Zhu, Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images, Field Crops Res. (2019), doi:10.1016/J.FCR.2019.02.022.
[8] L. Liu, H. Lu, Y. Li, Z. Cao, High-throughput Rice density estimation from transplantation to Tillering stages using deep networks, Plant Phenomics 2020 (2020) 1375957.
[9] C.Y. Yang, M.D. Yang, W.C. Tseng, Y.C. Hsu, G.S. Li, M.H. Lai, D.H. Wu, H.Y. Lu, Assessment of rice developmental stage using time series UAV imagery for variable irrigation management, Sensors 20 (2020) 5354, doi:10.3390/s20185354.
[10] H. Ge, F. Ma, Z. Li, Z. Tan, C. Du, Improved accuracy of phenological detection in rice breeding by using ensemble models of machine learning based on UAV-RGB imagery, Remote Sens. 13 (2021) 2678, doi:10.3390/rs13142678.
[11] X. Lu, J. Zhou, R. Yang, Z. Yan, Y. Lin, J. Jiao, F. Liu, Automated rice phenology stage mapping using UAV images and deep learning, Drones (2023), doi:10.3390/drones7020083.
[12] J. Qin, T. Hu, J. Yuan, Q. Liu, W. Wang, J. Liu, L. Guo, G. Song, Deep-learning-based rice phenological stage recognition, Remote Sens. 15 (2023) 2891, doi:10.3390/rs15112891.
[13] S.S. Sah, K.N.A. Maulud, S. Sharil, O.A. Karim, B. Pradhan, Monitoring of three stages of paddy growth using multispectral vegetation index derived from UAV images, Egypt. J. Remote Sens. Space Sci. 26 (2023) 989–998, doi:10.1016/j.ejrs.2023.11.005.