Data in Brief 61 (2025) 111843

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

InterDuPa-UAV: A UAV-based dataset for the classification of intercropped durian and papaya trees

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

[a] College of Engineering, Can Tho University, Can Tho, Vietnam
[b] CTU-AIMED Leading Research Team, Can Tho University, Can Tho, Vietnam
[c] Brubotics, Vrije Universiteit Brussel, Brussels, Belgium
[d] Flanders Make, Brussels, Belgium
[e] imec, Leuven, Belgium

ARTICLE INFO

Article history:
Received 18 April 2025
Revised 15 June 2025
Accepted 25 June 2025
Available online 1 July 2025

Dataset link: InterDuPa-UAV: A UAV-based Dataset for the Classification of Intercropped Durian and Papaya Trees (Original data)

Keywords:
Intercropped
Durian
Papaya
UAV
Image processing
Dataset
Machine learning

ABSTRACT

The growing use of Unmanned Aerial Vehicles (UAVs) in agriculture has made data collection more efficient and cost-effective, enabling the development of advanced solutions to enhance agricultural productivity. In this work, we present a dataset of intercropped durian (*durio zibethinus*) and papaya (*carica papaya*) trees, extracted from aerial images captured by a UAV. A total of 311 UAV-captured images were collected over a mixed plantation, where durian trees, being taller and broader, contrast visually with the shorter and more slender papaya trees. The trees were subsequently labeled into two categories based on type and visual characteristics: durian trees (3327 images) and papaya trees (2872 images). This dataset serves as a valuable resource for multiple tree species classification, spatial pattern analysis, and decision-making in precision agriculture using machine learning and deep learning methods. Insights derived from this dataset can support improved orchard management, including crop

* Corresponding author.
E-mail address: chlong@ctu.edu.vn (H.-L. Cao).
Social media: @HoangLongCao (H.-L. Cao)

https://doi.org/10.1016/j.dib.2025.111843

inventory, health monitoring, and optimization of intercropping strategies.

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
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 intercropped durian and papaya trees. The dataset comprises a total of 311 images of the mixed plantation. From these images, we extracted 3327 images of durian trees and 2872 images of papaya trees. The total size of the dataset is 3.8 GB.
Data source location	Institution: College of Engineering, Can Tho University Sampled paddy field: Chau Thanh, Hau Giang, Vietnam Latitude, longitude for collected data samples: 9°53'05.4"N 105°49'01"E Altitude: 25-30 meters
Data accessibility	Repository name: InterDuPa-UAV: A UAV-based Dataset for the Classification of Intercropped Durian and Papaya Trees Data identification number: https://doi.org/10.5281/zenodo. 15664908 Direct URL to data: https://zenodo.org/records/15664908 Instructions for accessing these data: Users can download directly via URL, extract, and use it.
Related research article	Luu, T.H., Phuc, P.N.K., Ngo, Q.H. and Nguyen, T.T., 2024. Drone Approach for Remote Sensing The Intercrop On Durian Plantations Using YOLOv5 Model. Science & Technology Asia, pp.128-137. https://doi.org/10.14456/scitechasia.2024.76

1. Value of the Data

- The dataset focuses on intercropped durian and papaya trees, distinguishing it from other datasets that may focus on monoculture plantations or different fruit types.
- The dataset was generated from aerial images captured using a UAV at altitudes of 25-30 meters, covering durian and papaya trees.
- Researchers can use this dataset to classify multiple tree species, analyze their spatial distribution, and support precision agriculture applications through machine learning and deep learning techniques.

2. Background

The adoption of Unmanned Aerial Vehicles (UAVs) in agriculture has accelerated the creation of high-resolution, scalable datasets for various applications, including crop monitoring, yield estimation, and species classification [1,2]. UAV-based data for plant classification have been used with diverse crops such as rice [3,4], maize [5,6], and orchards like citrus [7]. These data typically focus on monoculture systems, where the uniformity of plant types simplifies labeling, classification, and detection tasks.

Compared monoculture data, **fewer data focus on intercropped systems**, where two or more crops are cultivated simultaneously in the same field. For example, Hooshyar et al. collected UAV images from a mixed cultivation of lychee, longan, along with several other fruits [8]. Wu et al. conducted research on a diverse fruit data that included a wide range of species such as grapes,

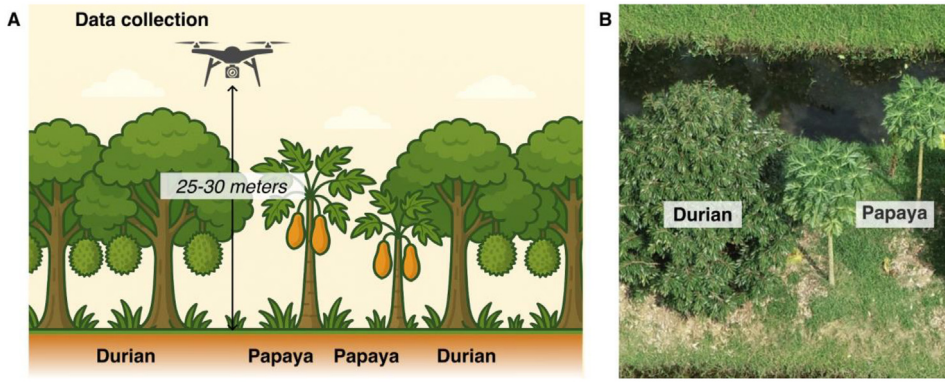

Fig. 1. Data collection was performed in an intercropped durian–papaya plantation using UAV technology. The durian trees are taller with dense canopies, while the papaya trees are shorter and more slender in structure.

peaches, apples, cherries, and others [9]. Intercropping, particularly with tree species, introduces challenges such as overlapping canopies, varying tree heights, and non- uniform spacing—all of which complicate aerial analysis. The lack of annotated datasets for such mixed-species plantations limits the development of models capable of handling complex agroforestry environments.

To address this gap, **we present a UAV-based dataset of intercropped durian (*Durio zibethinus*) and papaya (*Carica papaya*) trees**, collected in a real-world mixed plantation setting, see Fig. 1A. Durian and papaya differ in their physical structure—durian trees are taller with dense canopies and spiky fruits, while papaya trees are shorter with slender trunks and lobed leaves, Fig. 1B. Durian is a long-term crop that takes several years to mature, whereas papaya is a short-term crop with a faster growth cycle and earlier fruiting. This complementary growth pattern makes durian–papaya intercropping beneficial for maximizing land use efficiency, generating early income from papaya while waiting for durian trees to reach maturity. This type of durian–papaya intercropping is popular in Southeast Asian countries such as Vietnam and Thailand.

3. Data Description

Data was collected in Chau Thanh, Hau Giang province, Vietnam. The sampled crop is located at 9°53'05.4"N 105°49'01"E (Fig. 2). Our dataset consists of 311 high-resolution UAV images captured at different growth stages; and at 25 and 30 meters altitude. Each image are manu-

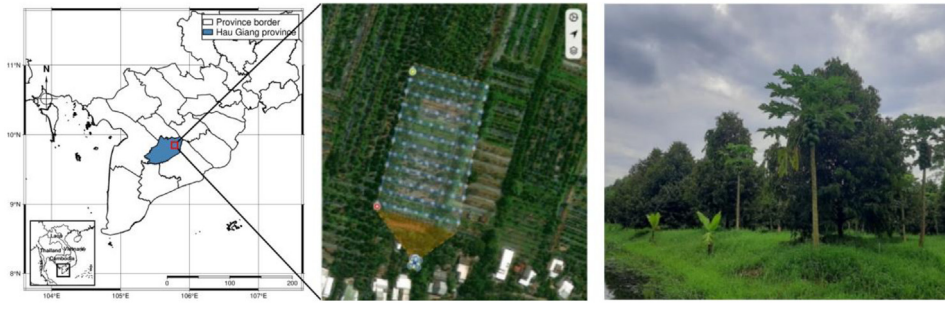

Fig. 2. Location of the sampled crop in Hau Giang province, Vietnam.

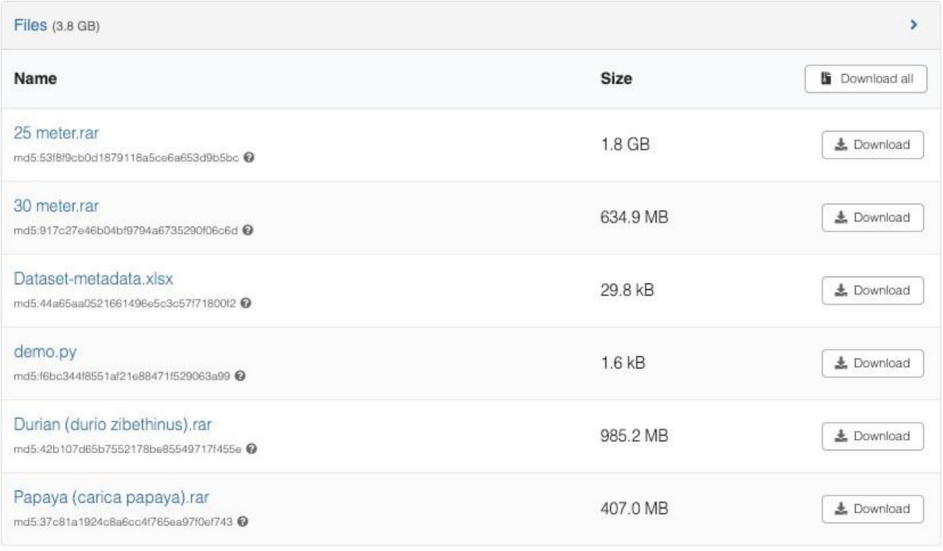

Fig. 3. Overview of the dataset files uploaded to Zenodo. The dataset includes UAV images captured at 25 and 30 meters altitude, as well as manually extracted individual images of durian (*Durio zibethinus*) and papaya (*Carica papaya*) trees. A structured metadata file and a basic Python demo script are included. The total dataset size is 3.8 GB.

ally cropped and annotated with tree types, supporting tasks such as tree species classification, canopy segmentation, and spatial distribution analysis. The dataset is available on Zenodo, with a total size of 3.8 GB, see Fig. 3 or an overview. We also included a structured metadata in the Zenodo repository to describe the entire dataset, along with a simple Python demo script for tree classification with a sample image using YOLO. Unlike existing datasets that focus on uniform plantations or low-level crops, this dataset offers a unique perspective on tree intercropping systems.

4. Experimental Design, Materials and Methods

4.1. Aerial data acquisition

We used a DJI-FC6310S camera mounted on a Phantom 4 Pro UAV to capture high-resolution images (4864 × 3648 pixels). The aerial images were collected in a mixed species plantation consisting of durian (*Durio zibethinus*) and papaya (*Carica papaya*) trees, at altitudes of 25- 30 meters. Data collection occurred on September 20, 2023, January 14, 2024, and January 20, 2024, at different growth stages, and under typical tropical field conditions. In total, 311 aerial images were acquired for analysis and annotation. Specific flight information and the number of images captured at each attitude are summarized in Table 1. Examples of raw images captured on the sample crop at 4 years and 5 years of growth are shown in Fig. 4.

Table 1
Flight information with UAV flight parameters used for data collection at two different altitudes.

Flight altitude	Number of images	Flight speed	Overlapped area
25 meters	231 images	2.5 m/s	60%
30 meters	80 images	3.0 m/s	60%

Fig. 4. Images of the intercropped durian and papaya trees at two growth stages: (A) 4 years of growth; (B) 5 years of growth.

4.2. Plant extraction from aerial images and image labelling

After acquiring a total of 311 high-resolution UAV images covering the mixed-species plantation area, a manual annotation and extraction process was carried out to isolate individual tree instances from the imagery. This process focused specifically on identifying and segmenting papaya and durian trees, the two primary species of interest in this study. Each tree was carefully examined and cropped from the original UAV images to ensure high-quality individual samples suitable for further analysis and model training. In total, we extracted and labeled 3327 individual durian tree images and 2872 individual papaya tree images, summarized in Table 2. A visual example of the extracted tree images is shown in Fig. 5.

Table 2
Summary of extracted individual plant instances from UAV imagery.

Tree species	Number of images
Durian (*durio zibethinus*)	3327
Papaya (*carica papaya*)	2872

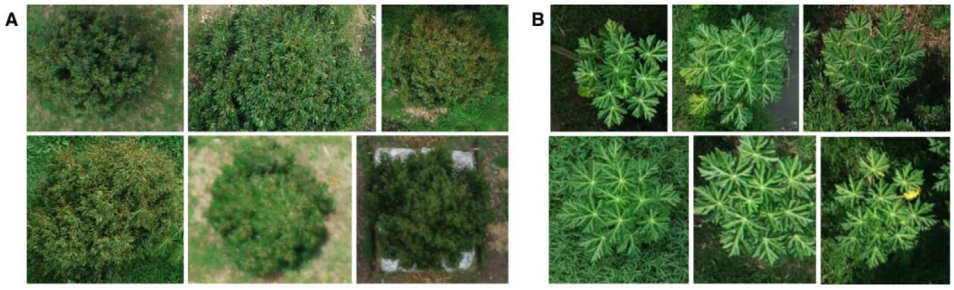

Fig. 5. Representative samples of manually extracted durian (A) and papaya (B) tree images from UAV data.

This dataset can be used to develop machine learning models that classify different tree species, study their spatial distribution, and support precision agriculture. As an example, we applied the YOLOv5 model to detect and locate papaya and durian trees using this dataset [10]. More advanced models can be trained to perform additional tasks, such as estimating canopy size, detecting growth anomalies, or predicting yield.

Limitations

The aerial images were collected during a single cropping season. Expanding the dataset to include images captured under different weather conditions and across more growing stages would increase its diversity and robustness.

Ethics Statement

The authors confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

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

Data Availability

InterDuPa-UAV: A UAV-based Dataset for the Classification of Intercropped Durian and Papaya Trees (Original data) (Zenodo).

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).

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] L. Epifani, A. Caruso, A survey on deep learning in uav imagery for precision agriculture and wild flora monitoring: Datasets, models and challenges, Smart Agric. Technol. (2024) 100625.
[2] S.A. Gade, M.J. Madolli, P. García-Caparrós, H. Ullah, S. Cha-um, A. Datta, S.K. Himanshu, Advancements in uav remote sensing for agricultural yield estimation: A systematic comprehensive review of platforms, sensors, and data analytics, Remote Sens. Applic. Soc. Environ. (2024) 101418.
[3] L. Wan, H. Cen, J. Zhu, J. Zhang, Y. Zhu, D. Sun, X. Du, L. Zhai, H. Weng, Y. Li, et al., Grain yield prediction of rice using multi-temporal uav-based rgb and multispectral images and model transfer–a case study of small farmlands in the south of china, Agric. For. Meteorol. 291 (2020) 108096.
[4] T.H. Luu, H.-L. Cao, Q.H. Ngo, T.T. Nguyen, I. El Makrini, B. Vanderborght, Rigad: An aerial dataset of rice seedlings for assessing germination rates and density, Data Brief. 57 (2024) 111118.
[5] F. Gnädinger, U. Schmidhalter, Digital counts of maize plants by unmanned aerial vehicles (uavs), Remote Sens. (Basel) 9 (6) (2017) 544.
[6] W. Chivasa, O. Mutanga, J. Burgueño, Uav-based high-throughput phenotyping to increase prediction and selection accuracy in maize varieties under artificial msv inoculation, Comput. Electron. Agric. 184 (2021) 106128.
[7] O. Csillik, J. Cherbini, R. Johnson, A. Lyons, M. Kelly, Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks, Drones 2 (4) (2018) 39.
[8] M. Hooshyar, Y.-S. Li, W.C. Tang, L.-W. Chen, Y.-M. Huang, Economic fruit trees recognition in hillsides: A cnn-based approach using enhanced uav imagery, IEEE Access, 2024.
[9] C. Wu, W. Jia, J. Yang, T. Zhang, A. Dai, H. Zhou, Economic fruit forest classification based on improved u-net model in uav multispectral imagery, Remote Sens. (Basel) 15 (10) (2023) 2500.
[10] T.H. Luu, P.N.K. Phuc, Q.H. Ngo, T.T. Nguyen, Drone approach for remote sensing the intercrop on durian plantations using yolov5 model, Sci. Technol. Asia (2024) 128–137.