



## Data Article

## Comprehensive dataset on ripening stages of strawberries and avocados: From unripe to rotten



Pooja Kamat<sup>a</sup>, Harsh Chandekar<sup>a</sup>, Lisanne Dlima<sup>a</sup>, Shilpa Gite<sup>a,b</sup>,  
Biswajeet Pradhan<sup>c,\*</sup>, Abdullah Alamri<sup>d</sup>

<sup>a</sup>Symbiosis Institute of Technology, Symbiosis International University, Near Lupin Research Park, Gram: Lavale, Tal: Mulshi, Pune, Maharashtra 412115, India

<sup>b</sup>Symbiosis Centre for Applied Artificial Intelligence (SCAAI), Symbiosis Institute of Technology, Symbiosis International University, Near Lupin Research Park, Gram: Lavale, Tal: Mulshi, Pune, Maharashtra 412115, India

<sup>c</sup>Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering & IT, University of Technology Sydney, NSW 2007 Australia

<sup>d</sup>Department of Geology and Geophysics, College of Science, King Saud University, Riyadh, Saudi Arabia

## ARTICLE INFO

## Article history:

Received 20 August 2024

Revised 23 February 2025

Accepted 8 May 2025

Available online 14 May 2025

Dataset link: [Comprehensive Dataset on Ripening Stages of Strawberries and Avocados: From Unripe to Rotten \(Original data\)](#)

## Keywords:

Strawberries

Avocados

Fruit maturity assessment

Precision agriculture

## ABSTRACT

This paper presents a novel and innovative 14,630 fruit images dataset, consisting of 1333 original images and the remaining augmented images for strawberry and avocado fruits. The dataset records the growth of strawberries and avocados in four different stages: unripe, partially ripe, ripe, and rotten. Though the fruit ripening process is commonly known, a lack of systematic datasets to show the fruit changing from an unripe state to a rotting state was prevalent for the two fruits in question. Over two months, the dataset was collected through rigorous tracking to effectively provide a measure of each of the fruits' conditions. The fruits were obtained from Mahabaleshwar farms in Maharashtra, India, as well as from local markets in Maharashtra and Pune. The fruits were monitored continuously from the time of harvesting, and all observed changes were carefully recorded. The uniqueness of this dataset is that it covers both strawberries and avocados, which have different patterns of ripening and are highly commercially valuable. The images were annotated using the online annotation tool - makesense.ai, with a

\* Corresponding author.

E-mail address: [Biswajeet.Pradhan@uts.edu.au](mailto:Biswajeet.Pradhan@uts.edu.au) (B. Pradhan).

total of 1499 bounding boxes for each fruit. By encompassing these two diverse fruit types, the dataset provides a valuable resource for researchers, agriculturalists, and food scientists to investigate and compare the ripening behaviours of different fruit species.

© 2025 The Author(s). Published by Elsevier Inc.  
This is an open access article under the CC BY license  
(<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Computer Vision and Pattern Recognition
Specific subject area	Fruit Maturity Detection
Type of data	Raw images with label files in .txt format)
Data collection	Photographs were regularly clicked from the camera with dual 12 MP with f/1.6, 26 mm (wide) and 1.4 µm pixel size for 25 days by placing the fruit (subject) on it from different angles in bright light conditions. For the strawberries, pictures were taken daily, from the flowering stage until the fruit reached the rotten stage. The photographic documentation was carried out during the day with abundant natural light and clear visibility, ensuring optimal image quality. Each strawberry was captured from approximately three different angles to provide comprehensive visual information for analysis. In the case of avocados, raw fruits were purchased from the market, and a controlled environment was set up using a white background for enabling consistent and uniform image acquisition..
Data source location	Dataset was collected from (i) Mahabaleshwar, Maharashtra, India; and (ii) Pune, Maharashtra, India.
Data accessibility	Repository name: mendeley.com Data identification number: 10.17632/zysvgmxcyz.1 Direct URL to data: <a href="https://data.mendeley.com/datasets/zysvgmxcyz/1">https://data.mendeley.com/datasets/zysvgmxcyz/1</a>
Related research article	

1. Value of the Data

- This dataset is a useful resource for machine learning solutions in fruit maturity detection and can contribute to the design of automated sorting and classification systems by ripeness stages.
- Food processing companies and agricultural scientists may utilize this data to improve post-harvest management, packaging choice, and shelf-life estimation.
- The data enables research in precision agriculture by providing a labeled and organized sequence of images spanning the entire range of fruit ripening, from non-ripe to spoiled.
- Knowledge of ripening may assist in minimizing food loss and enhancing composting planning, towards achieving sustainable principles in food science and agriculture.
- This data can be utilized to create consumer advisory systems using AI to find the best ripeness level for consumption, thereby ensuring maximum nutritional value.

2. Background

The research seeks to offer a comprehensive dataset to trace the evolution of avocados and strawberries from their unripe phases to almost degradation through recording the physical changes and appearance of both the fruits at every ripening day through meticulous observation and the compilation of images on a daily basis [1]. The research intends to contribute to our understanding of the fruit's ripening dynamics. Ultimately, an increased understanding of

fruit ripening will lead to improved models for ripening prediction, more effective post-harvest handling, and innovative fruit storage solutions for the forefront of modern food and agriculture [1]. With this dataset, researchers, horticultural professionals, and food scientists can study and compare how avocados and strawberries ripen, which should help advance our understanding of fruit ripening and also provide the necessary background to dig further into fruit physiology [2].

### 3. Data Description

Due to their considerable market value, the agriculture industry is particularly concerned with food waste, especially strawberries and avocados. Hence, these fruits were chosen for the dataset [3]. The dataset comprises 1333 images, each accompanied by corresponding text files and detailed bounding box information. The bounding box annotation uses the popular object-detection model YOLO (You-Only-Look-Once) [4]. The dataset includes four categories that describe the fruit's growth stages: semi-ripe, ripe, rotting, and unripe, resulting in a total of eight categories (Fig. 1). The dataset is arranged methodically and divided into two folders: one containing the photos and another labelled for labelled data storage. File names in these folders start with IMG\_0 and go up to IMG\_1332, totalling 1333 records in each fruit case. Many photos contain multiple bounding boxes, significantly enhancing the dataset's effectiveness. Across all the image collections combined, there are 1499 bounding boxes. These bounding boxes are crucial as they enable accurate localisation and clear identification of the specific developmental stages displayed by the avocados and strawberries in the images. Emphasising the importance of these bounding boxes is paramount [5]. They provide the structural basis for careful examination and in-depth study of the dynamic processes involved in the progressive ripening of avocados and strawberries. Their addition makes it possible to analyse how these fruits change over time.

To offer a thorough summary, the paper presents a tabular representation below illustrating the prevalence of various developmental stages within the split image categories (Table 1).

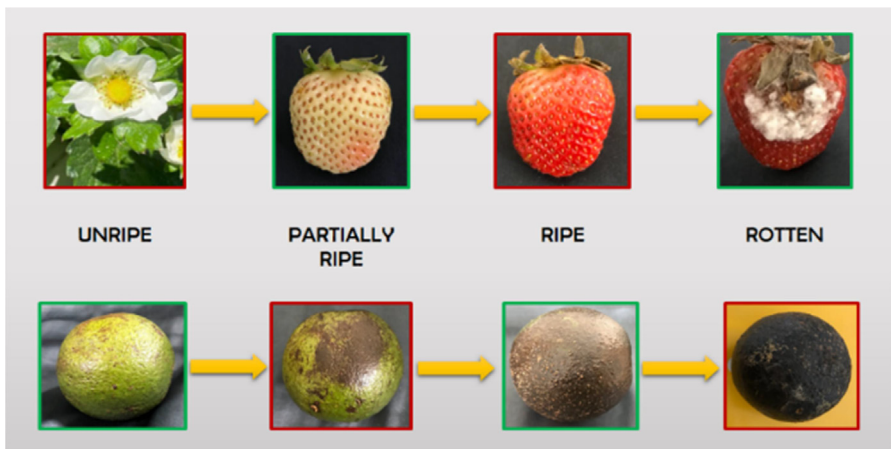


Fig. 1. Sample images of strawberry and avocado in different stages.

**Table 1**  
Category count of strawberries(S) and avocados (A).

Category	Count
SUnripe	114
SPripe	238
SRipe	189
SRotten	170
AUnripe	185
APripe	192
ARipe	129
ARotten	282



**Fig. 2.** Images before background change.

3.1. Dataset preprocessing

As seen in Fig. 2, the dataset utilised in this investigation consisted primarily of photos taken against a uniform background, which could have introduced bias by allowing the model to rely on background signals for fruit classification. To mitigate this bias, an important pre-processing step was replacing each image's background with a solid pastel color. Adopting shades of red, green, blue, purple, orange, and yellow aimed to enhance accurate fruit identification and prediction by reducing the impact of background-related information displayed in Fig 3.



**Fig. 3.** Images after background change.

**Table 2**

Balanced count of Strawberry and Avocado samples across different ripening stages.

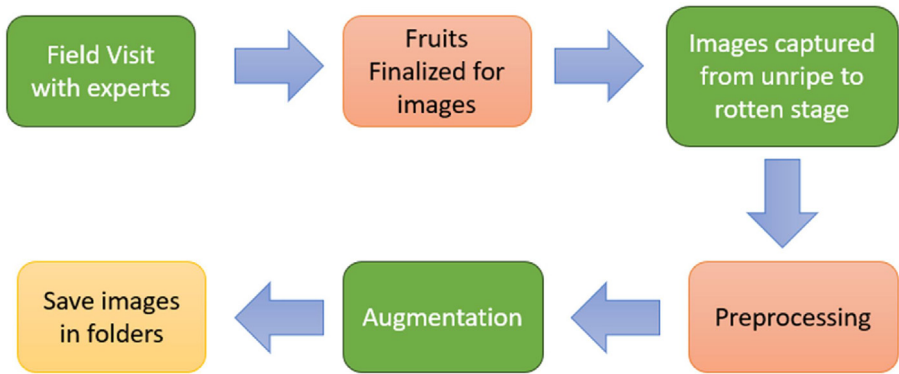
Category	Count
SUnripe	58
SPripe	58
SRipe	58
SRotten	58
AUnripe	58
APripe	58
ARipe	58
ARotten	58

### 3.2. Dataset balancing

The dataset balancing process was conducted to ensure a balanced representation across fruit categories. A subset of 58 images, along with their corresponding bounding boxes, was selected for each category, resulting in a total of 409 photos, in the final dataset due to this rigorous curation. The balanced dataset is stored in two folders, one holding the labels and the other the photos. It is important to note that an equal distribution of instances was obtained for each fruit type following this stringent balancing technique. Table 2 shows the precise count for each category.

## 4. Experimental Design, Materials and Methods

The experiment began with a thorough study of fruit taxonomy and ripening stages based on literature [6] and field studies. The subject fruits were monitored from their first stages to defoliation. Photographs were then collected and arranged subsequently in files named IMG\_0 → IMG\_1332. To facilitate model training and ensure color consistency across all images, the backgrounds were converted to pastel colors. The images were manually annotated to provide ground truth bounding boxes, which are essential for any object detection algorithm to function properly [7]. Augmentation techniques such as rotating images, carefully controlled blurring, and zooming



**Fig. 4.** Process of dataset preparation.

in/out were employed to enhance the dataset's comprehensiveness [8]. Further, this augmented dataset was then classified and assigned to systematic folders. Fig 4 shows the corresponding steps.

#### 4.1. Sampling and collection

- Strawberries: The dataset includes Sweet Charlie and Winter Dawn varieties, sourced from farms in Mahabaleshwar, Maharashtra, India.
- Avocados: The dataset focuses on the Hass variety, sourced from local vendors in Pune, Maharashtra, India.
- Data collection was conducted over 25 days, with daily monitoring of each fruit's ripening progression.

#### 4.2. Environmental conditions

- Strawberries were monitored in-field under natural daylight conditions, capturing multiple angles for comprehensive analysis.
- Avocados were placed in a controlled indoor environment, ensuring uniform lighting conditions and background standardization.
- Replications were performed through continuous monitoring, ensuring that each stage of ripeness was well-documented.

#### 4.3. Dataset annotations

There are a total of 4 categories for each fruit - unripe, partially ripe, ripe and rotten. The annotation tool makesense.ai was used to manually annotate all the images. Bounding boxes were created for annotation, and in some images, multiple bounding boxes were used due to the presence of several similar fruits within the same image frame. The images were then sequentially numbered and checked for any errors or anomalies.

#### 4.4. Dataset export

Upon completion of the annotation task, the entire dataset was resized to a resolution of  $1024 \times 1024$  pixels uniformly for the sake of data quality and consistency. The dataset has also

been shared on the Mendeley repository so that it is easily accessible and usable for future research and development.

## Limitations

The dataset, while being comprehensive, has a few limitations as follows:

- **Limited Geographic Scope:** Fruits were only obtained from two locations (Mahabaleshwar and Pune, Maharashtra, India), and they might not reflect the ripening trends of strawberries and avocados in other locations.
- **Limited Temporal Coverage:** Samples were taken within a span of two months, which could potentially fail to pick up the full scope of ripening behaviours over varying environmental conditions..
- **Focus on Two Fruit Types:** Although valuable, this dataset only includes information on strawberries and avocados, which may limit its applicability to generalizing the ripening processes of other fruit species.
- **Potential Bias in Fruit Selection:** Fruits were purchased from farms and local markets, which could introduce variability in ripening stages due to differences in handling, storage, and transportation practices.
- **Small Number of Original Images:** The dataset contains only 1333 original images, with the remaining images being augmented versions, potentially limiting the diversity of the dataset.

By addressing these limitations through further dataset expansion, the concept of fruit maturity detection can be significantly enhanced, leading to improved performance and broader applicability in real-world applications.

## Ethics Statement

The authors have confirmed that this study does not involve human subjects, animal experiments, or any data collected from social media platforms and follows the ethical requirement for publication in Data in Brief.

## CRediT Author Statement

**Pooja Kamat:** Conceptualization, Investigation, Data curation, Writing – original draft, Visualization; **Harsh Chandekar:** Conceptualization, Investigation, Data curation, Writing – original draft, Visualization; **Lissane Dlima:** Conceptualization, Investigation, Data curation, Writing – original draft, Visualization; **Shilpa Gite:** Writing – Review & Editing, Supervision; **Biswajeet Pradhan:** Writing – Review & Editing, Supervision, Funding, Resources; **Abdullah Alamri:** Writing – Review & Editing.

## Data Availability

[Comprehensive Dataset on Ripening Stages of Strawberries and Avocados: From Unripe to Rotten \(Original data\)](#) (Mendeley Data).

## Acknowledgements

This research was funded by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney.

## Declaration of Competing Interest

There are no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

## References

- [1] J. Giovannoni, C. Nguyen, B. Ampofo, S. Zhong, Z. Fei, The epigenome and transcriptional dynamics of fruit ripening, *Annu. Rev. Plant Biol.* 68 (2017) 61–84.
- [2] Anjali, A. Jena, A. Bamola, S. Mishra, I. Jain, N. Pathak, ... S. Akhtar, State-of-the-art non-destructive approaches for maturity index determination in fruits and vegetables: principles, applications, and future directions, *Food Prod. Process. Nutr.* 6 (1) (2024) 56.
- [3] I.K. Opara, U.L. Opara, J.A. Okolie, O.A. Fawole, Machine learning application in horticulture and prospects for predicting fresh produce losses and waste: a review, *Plants* 13 (9) (2024) 1200.
- [4] A. Vijayakumar, S. Vairavasundaram, Yolo-based object detection models: a review and its applications, in: *Multimedia Tools and Applications*, 2024, pp. 1–40.
- [5] M. Desai, H. Mewada, A novel approach for yoga pose estimation based on in-depth analysis of human body joint detection accuracy, *PeerJ Comput. Sci.* 9 (2023) e1152.
- [6] K.K. Saha, C. Weltzien, B. Bookhagen, M. Zude-Sasse, Chlorophyll content estimation and ripeness detection in tomato fruit based on NDVI from dual wavelength LiDAR point cloud data, *J. Food Eng.* (2024) 112218.
- [7] A. Behl, O. Hosseini Jafari, S. Karthik Mustikovela, H. Abu Alhaija, C. Rother, A. Geiger, Bounding boxes, segmentations and object coordinates: how important is recognition for 3d scene flow estimation in autonomous driving scenarios? in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2574–2583.
- [8] F. Jahangir, M.A. Khan, R. Damaševičius, F. Alblehai, A.I. Alzahrani, M. Shabaz, ... Y. Pandey, HGANet-23: a novel architecture for human gait analysis based on deep neural network and improved satin bowerbird optimization, in: *Signal, Image and Video Processing*, 2024, pp. 1–15.