



Data Article

FruitNet: Indian fruits image dataset with quality for machine learning applications



Vishal Meshram*, Kailas Patil*

Vishwakarma University, India

ARTICLE INFO

Article history:

Received 2 October 2021

Revised 2 December 2021

Accepted 3 December 2021

Available online 7 December 2021

Keywords:

Convolutional neural network

Computer vision

Deep learning

Fruit classification

Fruit detection

Fruit image dataset

Machine learning

ABSTRACT

Fast and precise fruit classification or recognition as per quality parameter is the unmet need of agriculture business. This is an open research problem, which always attracts researchers. Machine learning and deep learning techniques have shown very promising results for the classification and object detection problems. Neat and clean dataset is the elementary requirement to build accurate and robust machine learning models for the real-time environment. With this objective we have created an image dataset of Indian fruits with quality parameter which are highly consumed or exported. Accordingly, we have considered six fruits namely apple, banana, guava, lime, orange, and pomegranate to create a dataset. The dataset is divided into three folders (1) Good quality fruits (2) Bad quality fruits, and (3) Mixed quality fruits each consists of six fruits subfolders. Total 19,500+ images in the processed format are available in the dataset. We strongly believe that the proposed dataset is very helpful for training, testing and validation of fruit classification or reorganization machine learning model.

© 2021 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license
(<http://creativecommons.org/licenses/by/4.0/>)

* Corresponding authors.

E-mail addresses: vishal.meshram-020@vupune.ac.in (V. Meshram), kailas.patil@vupune.ac.in (K. Patil).

Social media:  (K. Patil)

Specifications Table

Subject	Machine learning, agriculture science, horticulture
Specific subject area	Fruits image dataset with quality classification (good, bad, and mixed)
Type of data	Indian fruits images
How data were acquired	Fruits images were using high resolution mobile phone camera in the natural and artificial light conditions with different backgrounds.
Data format	Raw
Parameters for data collection	The fruit dataset images are .jpg images of 256×256 dimension and resolution is 72 dpi.
Description of data collection	The fruits images were collected using high resolution mobile phones rear camera. The original .jpg images of fruits are of dimensions 3024×3024 . These images are resized to 256×256 dimensions. The dataset is categorized into 3 subfolders Good Quality Fruits, Bad Quality Fruits, and Mixed Quality Fruits. Further each folder contain six fruits classes namely Apple, Banana, Guava, Lime, Orange, Pomegranate. The images were taken at the different backgrounds and in different lighting conditions. The proposed dataset can be used for training, testing and validation of fruit classification or reorganization model.
Data source location	VISHWAKARMA UNIVERSITY Survey No. 2, 3, 4 Laxmi Nagar, Kondhwa Budruk, Pune - 411 048. Maharashtra, India. Latitude and longitude: 18.4603° N, 73.8836° E HUBTOWN COUNTRYWOODS SOCIETY Tilekar Nagar, Kondhwa Budruk, Pune - 411 048. Maharashtra, India. Latitude and longitude: 18.442866 ° N, 73.884894° E
Data accessibility	Repository name: FruitNet: Indian Fruits Dataset with quality (Good, Bad & Mixed quality) Data identification number(doi): 10.17632/b6fftwbr2v1 Direct URL to data: https://data.mendeley.com/datasets/b6fftwbr2v/1

Value of the Data

- The dataset is comprehensive which consist of 19500+ high-quality images of six different classes.
- The dataset consist of good quality, bad quality, and mixed quality fruit images.
- To the best of our knowledge this is the first open access dataset of indian fruits consists of good, bad and mixed quality fruits.
- This dataset is useful to build applications of fruit classification and detection with quality.
- The dataset will be useful for training, testing and validation of fruit classification or reorganization model.
- The dataset is useful to build fruit classification with quality applications which are beneficial for farmers, agriculture industries, wholesalers, hawkers, and customers, and fruit export companies.

1. Data Description

The profit percentage share of fruit market is substantial with respect to the total agriculture output [1–3]. In the agro-industry fast and accurate fruit classification is the highest need. The fruits can be classified into different classes as per their external features like shape, size and color using some computer vision and deep learning techniques [4–8]. The FruitNet dataset was created to include Indian fruits along with its quality parameters for those which are highly consumed or exported as per [9]. It consists of six classes of Indian fruits namely apple, banana, guava, lime, orange, and pomegranate. They further categorized into good quality, bad quality, and mixed quality. The fruit images were taken with different background, in different light

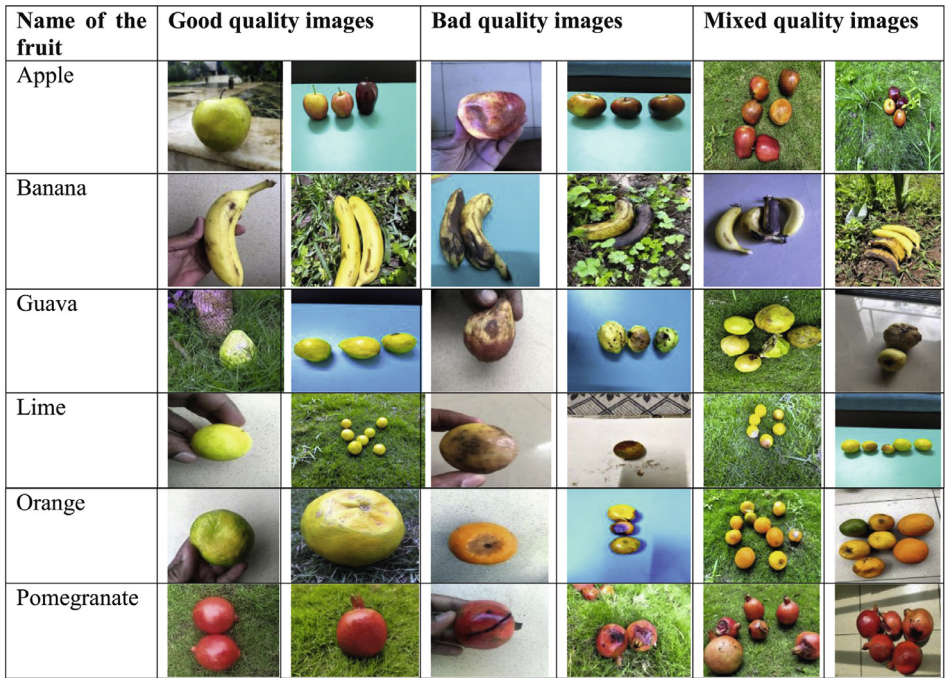


Fig. 1. Partial images of the dataset.

conditions in indoor and outdoor environment. The Fig. 1 shows the sample images in the dataset consisting of images taken in various environments.

2. Experimental Design, Materials and Methods

2.1. Experimental design

The image data acquisition process is shown in Fig. 2. The fruit images were acquired using three different make of camera's i.e. iPhone6 (Apple), ZUK (Z2 Plus), and Realme (Realme 5 Pro) mobile's high resolution rear camera. In all 19500+ images were captured using camera and then were segregated and saved in respective folders as per their quality and classification.

The data acquisition process steps are shown in Table 1. The fruit images are captured in the natural and artificial lighting conditions with different angles and background in months of July to October. Images pre-processing is done using python script. In the pre-processing we changed the dimensions to 256×256 which is standard resolution required to build object classification or object detection model.

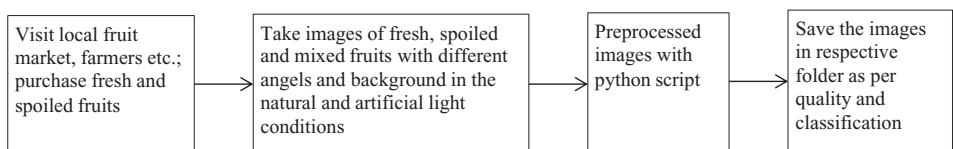


Fig. 2. Fruits data acquisition process.

Table 1

Data acquisition steps.

Sr. No.	Step	Duration	Activity
1.	Data Gathering	July to October	Daily captured the fruits images in the natural and artificial light with different angles and background.
2.	Pre-processing and creating dataset	November	Run the python script to pre-process the images (convert all images in 256 × 256 resolution) and save the images into respective folders as per their quality and classification (i.e. bad, good and mixed)

Table 2

Specification of image acquisition device.

Sr. No.	Camera Particulars	Details		
1	Camera maker	Apple	ZUK	Realme
2	Camera Model	iPhone 6	Z2 Plus	Realme 5 Pro
3	F-stop	f/2.2	f/2.2	f/1.8
4	Exposure time	1/25 s	1/214 s	1/33 s
5	ISO Speed	ISO-250	ISO-100	ISO-1120
6	Exposure bias	0 step	0 step	0 step
7	Focal length	4 mm	4 mm	5 mm
8	metering mode	Pattern	Centered Weighted Average	Unknown
9	Flash mode	No flash	No flash	No flash
10	35mm focal length	29	29	0

Table 3

Specification of images.

Sr. No.	Particulars	Details as per Fruit classes		
		Bad Fruit	Good Fruit	Mixed Fruit
1	Dimension	256 × 256	256 × 256	256 × 192
2	Width	256 pixels	256 pixels	256 pixels
3	Height	256 pixels	256 pixels	192 pixels
4	Horizontal Resolution	72 dpi	96 dpi	72 dpi
5	Vertical Resolution	72 dpi	96 dpi	72 dpi
6	Bit Depth	24	24	24
7	Resolution unit	2	2	2
8	Color representation	sRGB	sRGB	Uncalibrated

2.2. Materials or specification of image acquisition system

The fruit images are captured using Apple iPhone 6 with rear camera of 8 megapixels, Z2 Plus with rear camera of 13 megapixel, and Realme 5 Pro with rear camera of 48 megapixels. All dataset images of original size 3024 × 3024 were resized to 256 × 256 dimensions using a python script. The images are in .jpg images. The images acquired in variety of environmental conditions such as different light conditions, different background, and from different angles.

After capturing the images were organized as Bad quality, Good quality, and Mixed quality folders. Further each quality folder has six different folders of fruit classes i.e. apple, banana, guava, lime, orange, and pomegranate, respectively. The specifications of devices used for image acquisition and acquired images specifications are shown in Tables 2 and 3, respectively.

2.3. Method

All fruit images are acquired using three mobile make with a high resolution rear camera in different angles and different backgrounds. The original images of size 3024 × 3024 were resized

Table 4
FruitNet dataset details.

Quality classes	Fruit classes Considered	Image Taken in which Direction	Image Taken in different Backgrounds	No. of Images of each denomination	Total No. of Images
Bad quality	apple, banana, guava, lime, orange, pomegranate	Front Direction, Top View, Backward Direction, Bottom View, Direction Rotated 180 degrees,	Dark color, grass, light color, ground, multicolor	apple - 1141 banana - 1087 guava - 1129 lime - 1085 orange - 1159 pomegranate - 1187	6778
Good quality	apple, banana, guava, lime, orange, pomegranate	Front Direction, Top View, Backward Direction, Bottom View, Direction Rotated 180 degrees,	Dark color, grass, light color, ground, multicolor	apple - 1149 banana - 1113 guava - 1152 lime - 1094 orange - 1216 pomegranate - 5940	11664
Mixed quality	apple, banana, guava, lime, orange, pomegranate	Front Direction, Top View, Backward Direction, Bottom View, Direction Rotated 180 degrees,	Dark color, grass, light color, ground, multicolor	apple - 113 banana - 285 guava - 148 lime - 278 orange - 125 pomegranate - 125	1074
Total Number of Images in the Dataset					19526

to 256×256 using a python script. Table 4 describes the classes, number of image taken and the environments in which images are taken.

Ethics Statement

There is no funding for the present effort. There is no conflict of interest. The data is available in public domain.

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.

CRedit Author Statement

Vishal Meshram: Methodology, Software, Validation, Writing – original draft; **Kailas Patil:** Conceptualization, Supervision, Writing – review & editing.

References

- [1] A. Bhargava, A. Bansal, Fruits and vegetables quality evaluation using computer vision: a review, J. King Saud Univ. Comput. Inf. Sci. 33 (2021) 243–257, doi:[10.1016/j.jksuci.2018.06.002](https://doi.org/10.1016/j.jksuci.2018.06.002).
- [2] S. Behera, A. Rath, A. Mahapatra, P. Sethy, Identification, classification & grading of fruits using machine learning & computer intelligence: a review, J. Ambient Intell. Humaniz. Comput. (2020), doi:[10.1007/s12652-020-01865-8](https://doi.org/10.1007/s12652-020-01865-8).
- [3] V. Meshram, K. Patil, V. Meshram, D. Hanchate, S. Ramteke, Machine learning in agriculture domain: a state-of-art survey, Artif. Intell. Life Sci. (2021) Article 100010, doi:[10.1016/j.aillsci.2021.100010](https://doi.org/10.1016/j.aillsci.2021.100010).
- [4] X. Chen, G. Zhou, A. Chen, L. Pu, W. Chen, The fruit classification algorithm based on the multi-optimization convolutional neural network, Multimed. Tools Appl. 80 (2021) 11313–11330, doi:[10.1007/s11042-020-10406-6](https://doi.org/10.1007/s11042-020-10406-6).

- [5] S. Behera, A. Rath, P. Sethy, Maturity status classification of papaya fruits based on machine learning and transfer learning approach, *Inf. Process. Agric.* 8 (2021) 244–250, doi:[10.1016/j.inpa.2020.05.003](https://doi.org/10.1016/j.inpa.2020.05.003).
- [6] H. Ayaz, E. Rodríguez-Esparza, M. Ahmad, D. Oliva, M. Pérez-Cisneros, R. Sarkar, Classification of apple disease based on non-linear deep features, *Appl. Sci.* 11 (2021) 6422, doi:[10.3390/app11146422](https://doi.org/10.3390/app11146422).
- [7] M. Momeny, A. Jahanbakhsh, K. Jafarnejad, Yu-Dong Zhang, Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach, *Postharvest Biol. Technol.* 166 (2020) 111204, doi:[10.1016/j.postharvbio.2020.111204](https://doi.org/10.1016/j.postharvbio.2020.111204).
- [8] V.A. Meshram, K. Patil, S.D. Ramteke, MNet: a framework to reduce fruit image misclassification, *Ingénierie des Syst. d'Inf.* 26 (2) (2021) 159–170, doi:[10.18280/isi.260203](https://doi.org/10.18280/isi.260203).
- [9] https://apeda.gov.in/apedawebsite/six_head_product/FFV.htm.