Data in Brief 38 (2021) 107312

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Dataset for Hop varieties classification

Pedro Castro, Eduardo Luz*, Gladston Moreira

Computing Department, Federal University of Ouro Preto, Ouro Preto-MG 35400-000, Brazil

ARTICLE INFO

Article history:
Received 20 July 2021
Revised 14 August 2021
Accepted 19 August 2021
Available online 24 August 2021

Keywords:
Hop varieties
Plant recognition
Leaf recognition

ABSTRACT

Humulus lupulus L., also known as hops, is a vine whose flowers are a major component in brewing. It delivers flavor, bitterness, and aroma to beer and also aids in foam stabilization. Furthermore, it plays an important role in beer conservation due to its antimicrobial and antioxidant properties, which have recently been studied for food preservation. Hops can also be found in the production of cosmetics and is considered healthy food.

There are more than 250 cataloged varieties of hops, and among the main attributes that differ from each other are alpha-acids, beta-acids, and essential oils. Those components give the beer a unique combination of characteristics, and may even influence its category. There are many ways to identify the hop variety from its acids and essential oils using methods such as chromatography, mass spectrometry, capillary electrophoresis, and nuclear magnetic resonance. However, these methods demand expensive and complex equipment, inaccessible or unavailable to most beer producers.

In this work, we present a database that includes 1592 images of hop leaves, from 12 popular hop varieties in southeastern Brazil. From these images, it is possible to explore methods of pattern recognition and machine learning to classify hop varieties

* Corresponding author.
E-mail address: eduluz@ufop.edu.br (E. Luz).
Social media: (E. Luz), (G. Moreira)

https://doi.org/10.1016/j.dib.2021.107312

Specifications Table

Subject	Agricultural Sciences
Specific subject area	Plant variety classification
Type of data	Figure, text annotation
How data were acquired	Cellphone camera sensor Instruments: Motorola Moto G7, Samsung Galaxy A11 and Apple iPhone 11
Data format	JPG, XML (Pascal VOC XML Annotation Format for label and bounding box)
Parameters for data collection	The only constraint imposed on data collection was to include at least one entire hop leaf.
Description of data collection	Images were taken in a natural environment, with varying climate, light, focus, occlusion, resolution, distance, and angle. Samples contain adult and young leaves. There was no removal of outliers or low-quality images. No pre-processing was adopted either.
Data source location	Institution: Atlântica Hops City/Town/Region: Juquiá, São Paulo / Litoral Sul Paulista Country: Brazil Institution: Hops Brasil City/Town/Region: Cedral, São Paulo / São José do Rio Preto Country: Brazil Institution: Brazuca Lúpulos City/Town/Region: Petrópolis, Rio de Janeiro / Serrana Country: Brazil
Data accessibility	https://doi.org/10.6084/m9.figshare.14933178

Value of the Data

- Classify the hop variety is of paramount importance for brewers [1,2] and also other applications (cosmetic, medicinal) [3–5]. Current methods are expensive and complex [6–9]. Computer vision may be a viable path.
- The data can provide ground reference for testing and validating machine learning methods to support image classification of hop varieties. Brewers could apply these methods and try to improve the production process by better-controlling hop variety.
- This dataset is the first hop variety dataset publicly available.

1. Data Description

The UFOP Hop Varieties Dataset (UFOP-HVD) consists of 1592 images of hop leaves captured in 3 plant nurseries in Brazil. The leaf images were acquired by different people and different mobile devices (camera sensors) in order to increase the representativeness of the database. Images were taken in a natural environment, with varying climate, light, focus, occlusion, resolution, distance, and angle. Samples contain adult and young leaves. There was no removal of outliers or low-quality images. No pre-processing was adopted either. There are more than 250 cataloged varieties [10] of this plant and, among the main attributes that differ from each other, are the alpha-acids, beta-acids, and [11] essential oils. The present dataset focuses on 12 varieties popular in southeastern Brazil. Fig. 1 contains examples of each of the 12 varieties used in this dataset. Table 1 contains the number of images of each class as well as the image resolutions according to the devices (Motorola Moto G7, Samsung Galaxy A11, and Apple iPhone 11).

Each image may contain one or more leaves of a hop of the same variety. All leaves were marked with *bounding boxes* as shown in the examples in Fig. 2 and labeled by field specialists. The leaf with the largest area was labeled as main (*bounding box* red), while the others as extra leaves (in yellow). The bounding box annotations are provided with the *dataset* in the Pascal VOC XML format [12].

Table 1
Hop variety distribution per image resolution.

	Resolution (pixel)								
Hop Variety	1040 × 520	1032 × 581	3391 × 2345	4096 × 2304	3683 × 3024	3024 × 4032	4032 × 3024	3072 × 4096	4096 × 3072
Cascade	51	9	0	0	1	3	46	0	0
Centennial	0	0	0	0	0	0	0	83	42
Cluster	30	0	0	23	0	0	50	0	0
Comet	55	10	0	0	0	3	47	12	44
Hallertau Mittelfrueh	0	0	0	0	0	0	0	67	74
Nugget	0	8	0	58	0	2	48	0	0
Saaz	0	7	1	0	0	1	48	39	76
Sorachi Ace	0	8	0	0	0	0	50	76	59
Tahoma	0	8	0	0	0	3	47	31	30
Triple Pearl	60	8	0	0	0	1	49	0	0
Triumph	0	9	0	42	0	5	45	0	0
Zeus	0	8	0	0	0	8	42	63	2
Total	**196**	**75**	**1**	**123**	**1**	**26**	**472**	**371**	**327**

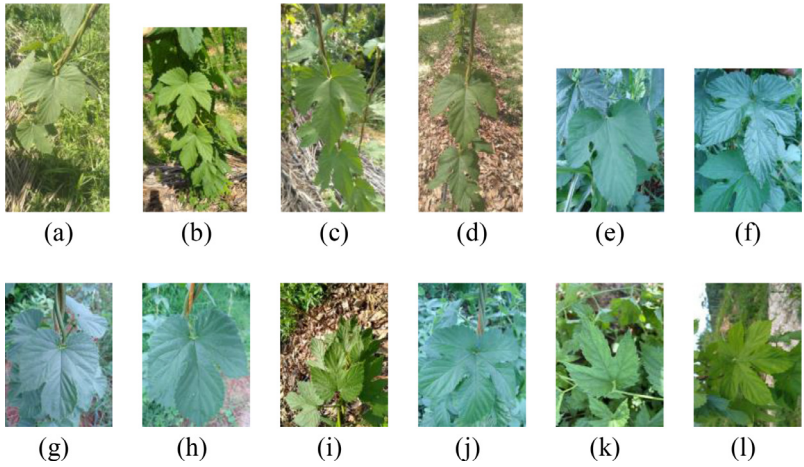

Fig. 1. Examples of the 12 Hops varieties contained in this data set: (a) Cascade; (b) Nugget; (c) Cluster; (d) Triple Pearl; (e) Hallertau Mittelfrueh; (f) Centennia; (g) Saaz; (h) Sorachi Ace; (i) Tahoma; (j) Comet; (k) Triumph; (l) Zeus.

Fig. 2. Labeling methodology: leaf with the largest area was labeled as main (red bounding box), while the others as extra leaves (yellow).

See below an example of an XML file available with the dataset:

```
<annotation>
  <folder>cascade</folder>
  <filename>cascade_l1_13.jpg</filename>
  <path>../validation/cascade/cascade_l1_13.jpg</path>
  <source>
    <database>UFOP-HVD</database>
  </source>
  <size>
    <width>520</width>
    <height>1040</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>cascade</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>225</xmin>
      <ymin>2</ymin>
      <xmax>475</xmax>
      <ymax>235</ymax>
    </bndbox>
  </object>
  <object>
    <name>cascade</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>1</xmin>
      <ymin>2</ymin>
      <xmax>238</xmax>
      <ymax>257</ymax>
    </bndbox>
  </object>
  <object>
    <name>cascade</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>102</xmin>
      <ymin>325</ymin>
      <xmax>391</xmax>
      <ymax>622</ymax>
    </bndbox>
  </object>
  <object>
    <name>cascade</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>291</xmin>
      <ymin>580</ymin>
      <xmax>428</xmax>
      <ymax>702</ymax>
    </bndbox>
  </object>
  <object>
    <name>cascade</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>318</xmin>
      <ymin>717</ymin>
      <xmax>472</xmax>
      <ymax>890</ymax>
    </bndbox>
  </object>
</annotation>
```

We emphasize that each XML refers to an acquired image and that it can contain more than one leaf and consequently more than one bounding box annotation. The XML presented corresponds to the left image of Fig. 2.

From the original *dataset*, a new one was generated containing all the leaves cut by the *bounding boxes*. This other set of data is called Cropped Dataset. The number of leaves per class in this database and the average of extracted leaves per image are displayed in Table 2.

Table 2
Cropped Dataset description.

Hop Variety	Number average extracted leaf / image	Number total leaf / variety
Cascade	**3,76**	**414**
Centennial	**1,62**	**202**
Cluster	**4,36**	**449**
Comet	**1,88**	**321**
Hallertau Mittelfrueh	**1,65**	**232**
Nugget	**3,81**	**442**
Saaz	**2,05**	**353**
Sorachi Ace	**2,02**	**390**
Tahoma	**2,30**	**274**
Triple Pearl	**3,53**	**417**
Triumph	**3,77**	**381**
Zeus	**2,31**	**284**
Total	**2,61**	**4159**

2. Experimental Design, Materials and Methods

The database is divided into three sets of data (70% for training, 15% for validation and 15% for test) in order to standardize the evaluation of machine learning methods. The division can be seen in Tables 3 and 4.

Table 3
Number of images of each class for each of the partitions (train, validation, test), in dataset without pre-processing.

Hop Variety	# Train	# Validation	# Test	Total
Cascade	78	16	16	**110**
Centennial	89	18	18	**125**
Cluster	73	15	15	**103**
Comet	121	25	25	**171**
Hallertau Mittelfrueh	99	21	21	**141**
Nugget	82	17	17	**116**
Saaz	122	25	25	**172**
Sorachi Ace	137	28	28	**193**
Tahoma	85	17	17	**119**
Triple Pearl	84	17	17	**118**
Triumph	71	15	15	**101**
Zeus	87	18	18	**123**
Total	**1128**	**232**	**232**	**1592**

Since the database is unbalanced in terms of images per class, results must be reported using recall, precision and F1-score per class.

Table 4
Number of leaves of each class for each of the partitions (train, validation, test), in Cropped Dataset.

Hop Variety	# Train	# Validation	# Test	Total
Cascade	318	39	57	**414**
Centennial	158	23	21	**202**
Cluster	316	58	75	**449**
Comet	212	52	57	**321**
Hallertau Mittelfrueh	165	30	37	**232**
Nugget	301	74	67	**442**
Saaz	265	40	48	**353**
Sorachi Ace	286	43	61	**390**
Tahoma	203	32	39	**274**
Triple Pearl	299	56	62	**417**
Triumph	260	56	65	**381**
Zeus	196	48	40	**284**
Total	**2979**	**551**	**629**	**4159**

Ethics Statement

Not applicable.

CRediT Author Statement

Pedro Castro: Conceptualization, Data curation, Writing–original draft; **Eduardo Luz**: Conceptualization, Experimental Design, Writing-Reviewing and editing; **Gladston Moreira**: Conceptualization, Writing-Reviewing and editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Data Availability

UFOP Hop Varieties Dataset (UFOP-HVD) (Original data) (figshare).

Acknowledgments

The authors would like to thank Atlântica Hops, Hops Brasil and Brazuca Lúpulos for providing the images, Lúpulo Zona da Mata for technical contribution and UFOP for their financial support.

References

[1] G. Astray, P. Gullón, B. Gullón, P.E. Munekata, J.M. Lorenzo, Humulus lupulus l. as a natural source of functional biomolecules, Appl. Sci. 10 (2020) 5074.

[2] C. Almaguer, C. Schönberger, M. Gastl, E.K. Arendt, T. Becker, Humulus lupulus–a story that begs to be told. a review, J. Inst. Brew. 120 (2014) 289–314.
[3] T. Nuutinen, Medicinal properties of terpenes found in cannabis sativa and humulus lupulus, Eur. J. Med. Chem. 157 (2018) 198–228.
[4] L. Chadwick, G. Pauli, N. Farnsworth, The pharmacognosy of humulus lupulus l.(hops) with an emphasis on estrogenic properties, Phytomedicine 13 (2006) 119–131.
[5] P. Zanoli, M. Zavatti, Pharmacognostic and pharmacological profile of humulus lupulus l, J. ethnopharmacol. 116 (2008) 383–396.
[6] M. Kovačevič, M. Kač, Determination and verification of hop varieties by analysis of essential oils, Food Chem. 77 (2002) 489–494.
[7] R.A. Shellie, S.D. Poynter, J. Li, J.L. Gathercole, S.P. Whittock, A. Koutoulis, Varietal characterization of hop (humulus lupulus l.) by gc–ms analysis of hop cone extracts, J. Sep. Sci. 32 (2009) 3720–3725.
[8] M.A. Farag, E.A. Mahrous, T. Lübken, A. Porzel, L. Wessjohann, Classification of commercial cultivars of humulus lupulus l.(hop) by chemometric pixel analysis of two dimensional nuclear magnetic resonance spectra, Metabolomics 10 (2014) 21–32.
[9] L.M. Duarte, L.H.C. Adriano, M.A.L. de Oliveira, Capillary electrophoresis in association with chemometrics approach for bitterness hop (humulus lupulus l.) classification, Electrophoresis 39 (2018) 1399–1409.
[10] J. Healey, The Hops List: 265 Beer Hop Varieties From Around the World, Julian Healey, 2016.
[11] B. Steenackers, L. De Cooman, D. De Vos, Chemical transformations of characteristic hop secondary metabolites in relation to beer properties and the brewing process: a review, Food Chem. 172 (2015) 742–756.
[12] M. Everingham, J. Winn, The pascal visual object classes challenge 2012 (voc2012) development kit, Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep 8 (2011) 5.