

# 中国科学数据



CHINA SCIENTIFIC DATA 第7卷 第1期 2022年3月(季刊)

专题：国家科学数据中心联合专刊



<http://www.csdata.org/p/issue/549/>

《中国科学数据（中英文网络版）》2022年第7卷第1期

## 目 录

### 国家科学数据中心联合专刊

《国家科学数据中心联合专刊》总序

苏靖

《国家科学数据中心联合专刊》子序

胡良霖

生态系统要素长期观测（EcoLTO）数据产品规范研制

郭学兵\*，唐新斋，苏文，何洪林

生物数据的标准化与微生物数据标准的发展

孙定中\*，马俊才\*

2020年基于VLF/LF三维闪电定位系统的全国闪电数据集

孟晓阳，王佳权，马启明\*，苑尚博，宋佳军，周晓，肖芳，王亚惠

2007–2020中国探月工程VLBI测量数据集

陈中\*，郑为民，陈肖，薛岩松

1960年武汉站电离层测高仪数据集

赵秀宽\*，宁百齐，胡连欢，刘立波，李国主，解海永，李凤琴，杨敏

社会公用计量标准数据集

田锋\*，智峰，赵若凡

中药材化学成分的晶体结构数据集

李英勇

## IDADP-葡萄病害识别研究图像数据集

袁媛\*, 陈雷\*

## 基于动物志的粉蝶形态特征数据集

李文杰, 王江宁, 卜翠萍, 葛斯琴, 林聪田, 韩艳, 纪力强\*

## 1982–2020 年中国 5 km 分辨率逐月 NDVI 数据集

徐洋, 杨雅萍\*

## 1992–2018 年中国及其毗邻地区基于植被功能类型的土地覆盖与香农多样性指数数据集

严涛, 金佳鑫\*, 朱青松, 刘颖

## 黄河流域 250 m 分辨率植被生长季时空演变数据集 (2000–2020 年)

白燕\*, 杨雅萍, 孙九林

## 三北工程区生态系统土壤保持能力评估数据集 (2000–2020 年)

张乃静, 肖云丹, 侯瑞霞, 魏胜蓉, 纪平\*

## 基于水资源承载力的华北地区降水与地下水要素数据集 (2005–2016 年)

肖云丹, 张乃静, 王俊荣, 侯瑞霞, 魏胜蓉, 纪平\*

## 2015 年长三角地区 30 m 土地覆被融合数据

陈逸聪, 邵华, 李杨\*, 戴玲

## 云南漾濞石门关景区动植物资源数据集

肖翠, 金冬梅, 李颖超, 李晓京, 张路杨, 郑柏岩, 吉小冬, 林秦文\*

## 基于“两叶”模型的 2001–2016 年贵州省 LAI 与 APAR 数据集

刘颖, 周士杰, 金佳鑫, 严涛

## 地球大数据科学工程专题

### 中国城市 2020 年建成区数据集

孙洁, 孙中昶\*, 郭华东, 王建波, 蒋会平, 高建

## 中国生态系统研究网络 (CERN) 专题

### 2011–2020 年云南元江干热河谷生态站气象监测数据集

孙晨娜, 杨大新, 宋清海\*, 陈爱国\*, 闻国静, 张树斌, 张晶, 段兴武, 金艳强

## 其他数据论文

### MuSyQ 高分 16 米分辨率 10 天合成的叶面积指数产品 (2018–2020 年中国 01 版)

张虎, 李静\*, 张召星, 吴善龙, 仲波, 柳钦火

### MuSyQ 高分 16 米空间分辨率 10 天合成的植被覆盖度产品 (2018–2020 年中国 01 版)

赵静, 李静\*, 张召星, 吴善龙, 仲波, 柳钦火\*

### MuSyQ 高分 16 米空间分辨率 10 天合成的 NDVI 植被指数产品 (2018–2020 年中国 01 版)

李松泽, 李静\*, 于文涛, 张召星, 吴善龙, 仲波, 柳钦火

### MuSyQ 高分 30 米空间分辨率 10 天合成的叶片叶绿素含量产品 (2019–2020 年中国 01 版)

李静\*, 张虎, 王晓函, 张召星, 谷晨鹏, 文远, 褚天嘉, 柳钦火

### 2015–2020 年广东省沿岸红树林矢量数据集

刘叶取, 张立\*, 郭康丽, 党二莎, 唐世林

## IDADP-葡萄病害识别研究图像数据集

ISSN 2096-2223

CN 11-6035/N



文献 DOI:

10.11922/11-6035.csd.2021.0077.zh

数据 DOI:

10.11922/sciencedb.j00001.00311

文献分类: 信息科学

收稿日期: 2021-11-09

开放同评: 2021-11-29

录用日期: 2022-03-19

发表日期: 2022-03-30

袁媛<sup>1,2\*</sup>, 陈雷<sup>1,2\*</sup>

1. 中国科学院合肥物质科学研究院, 智能机械研究所, 合肥 230031

2. 国家基础学科公共科学数据中心, 北京 100190

**摘要:** 构建可供机器学习建模使用的图像数据集对于实现高效的农业病虫害识别至关重要。本数据集收集整理了白粉病、霜霉病等7种葡萄病害图像数据, 共计3622张, 约697 MB。本数据集可为葡萄病害识别研究领域提供宝贵的基础数据资源, 同时可作为大数据环境下机器学习建模的标准图库, 对促进农业病害图像识别研究的发展具有重要的实际应用价值。

**关键词:** IDADP; 葡萄病害; 图像识别; 机器学习; 训练样本

### 数据库(集)基本信息简介

数据库(集)名称	IDADP-葡萄病害识别研究图像数据集
数据作者	袁媛、陈雷
数据通信作者	袁媛(yuanyuan@iim.ac.cn); 陈雷(chenlei@iim.ac.cn)
数据时间范围	2018–2021年
地理区域	中国境内
数据量	697 MB
数据格式	*.JPG
基金项目	国家自然科学基金面上项目(31871521、32071901); 国家基础学科公共科学数据中心“面向农业图像理解的机器学习数据集”(NBSDC-DB-20)。
数据服务系统网址	<a href="http://www.icgroupcas.cn/website_bchtk/tuku_putao.html">http://www.icgroupcas.cn/website_bchtk/tuku_putao.html</a> <a href="http://www.doi.org/10.11922/sciencedb.j00001.00311">http://www.doi.org/10.11922/sciencedb.j00001.00311</a>
数据库(集)组成	本数据集共有7种葡萄病害图像, 每种病害对应一个文件夹, 包括葡萄白粉病、葡萄花叶病毒病、葡萄黑霉病、葡萄灰霉病、葡萄溃疡病、葡萄霜霉病和葡萄酸腐病。每个文件夹中包含该病害图像以流水号命名的原始JPG文件, 以及介绍该病害基本信息与防治方法的intro.txt文件。本数据集共有JPG图像3622张。

葡萄是重要的经济作物, 在世界多个国家和地区广泛种植。我国葡萄栽培面积大且产量较高, 据国家统计局2019年的数据显示, 仅葡萄园种植面积就有72.62万公顷, 年产量高达1419.54万吨。葡萄病虫害是直接影响葡萄产量和品质的关键因素。随着栽培方式的改变和栽培面积的扩大, 葡萄病虫害种类逐渐增多, 病虫害的威胁逐渐加大<sup>[1]</sup>。

\* 论文通信作者

袁媛: yuanyuan@iim.ac.cn

陈雷: chenlei@iim.ac.cn



人工智能、大数据和计算机视觉技术的发展为农作物病虫害图像识别提供了新方法。尤其是 2006 年 Hinton 提出深度学习<sup>[2]</sup>以来，在诸多领域尤其是图像分类方面取得了显著的效果。然而，深度神经网络方法的使用需要有较大规模的训练数据，其数据的规模和质量在很大程度上决定了训练的效果。葡萄病害图像识别需要大规模的葡萄病害图像基础数据。文章作者在有关项目的支持下，通过采集、整合数据，建设了农业病虫害研究图库 (IDADP)<sup>[3]</sup>，涵盖农业病虫害图像采集、分类、标记、存储与建模等多方面的内容，面向科研学者与农技人员两大类用户群体提供农业病害在线诊断及相关的技术咨询等服务。此次发布的数据集是 IDADP 中的葡萄病害图像资源，共计 7 种病害 3622 张图像，可为机器学习建模提供训练和测试样本。

由于 IDADP-葡萄数据集和《大田作物病害识别研究图像数据集》<sup>[4]</sup>的数据采集和处理方法、数据存储和管理、数据质量控制和评估、使用方法等都相同，因此这些部分请参考关联文献[4]，具体不再赘述，两者的比较如表 1 所示。两个数据集都属于 IDADP，由于病害种类多，数据采集及整理需要时间，因此仅当数据集满足深度神经网络建模的需求时发布。

表 1 数据集比较

Table 1 Dataset comparison

比较项目	说明
数据采集和处理方法	图像采集设备增加了佳能 EF-M 28 F/3.5 Macro IS STM 微距镜头
数据样本描述	相同
数据质量控制和评估	葡萄病害样本图像来源主要是拍摄各类生产基地自然发生的病害图像
数据价值	面向的作物种类不同
数据使用方法	相同

数据集依然以图像数据库形式进行存储与管理，此处以葡萄病害数据为例作为示范，具体如表 2 所示。本数据集中的部分图像样本示例如图 1 所示。

表 2 葡萄病害图像数据词典的数据表字段与示例

Table 2 Data table fields and examples for the grape disease image data dictionary

列 1	列 2	列 3	列 4	列 5
作物名称	器官名称	病虫害名称	图像路径	备注
葡萄	叶部, 果实	霜霉病	葡萄霜霉病\IMG3283.jpg	-
葡萄	叶部, 果实	白粉病	葡萄白粉病\IMG2106.jpg	-
葡萄	果实	溃疡病	葡萄溃疡病\IMG2007.jpg	-

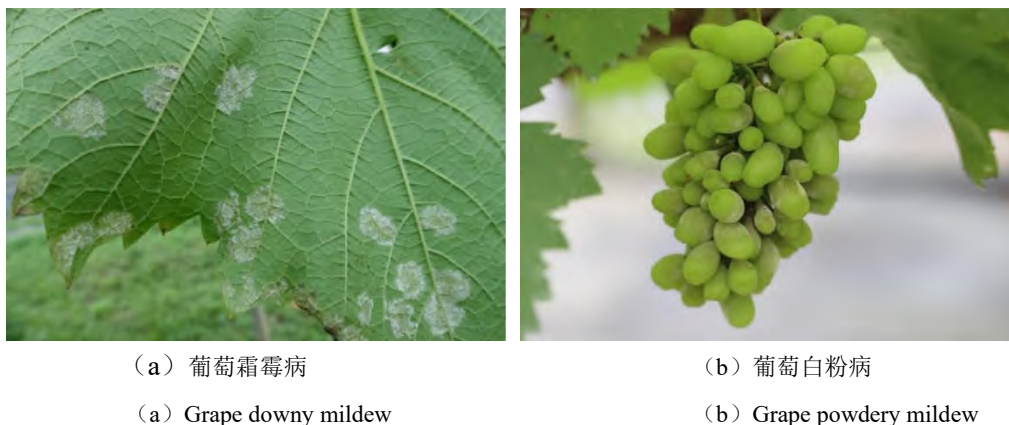


图1 葡萄病害识别研究图像数据集中的样本示例

Figure 1 Examples of the grape disease image dataset

由于数据版权原因，本文所上传的图像数据已进行适当压缩，如需获取更高质量的图像数据请前往图库网站<sup>[3]</sup>或联系作者。

## 致 谢

感谢安徽省农业科学院黄亮、陆丽娟提供葡萄病害图像拍摄基地，并协助对葡萄病害图像进行鉴定与分类。

## 数据作者分工职责

袁媛（1981—），女，安徽省肥东县人，博士，副研究员，研究方向为计算机视觉方法及在农业病害图像识别中的应用。主要承担工作：本数据集原始数据的采集、整理与相关软件工具的研发。

陈雷（1981—），男，安徽省巢湖市人，博士，副研究员，研究方向为机器学习理论方法及在大数据环境下的应用。主要承担工作：本数据集平台的规划、建设和维护。

## 参考文献

- [1] 段长青, 刘崇怀, 刘凤之, 等. 新中国果树科学研究 70 年: 葡萄[J]. 果树学报, 2019, 36(10): 1292–1301. DOI:10.13925/j.cnki.gsxz.2019.10.1292. [DUAN C Q, LIU C H, LIU F Z, et al. Fruit scientific research in New China in the past 70 years: grape[J]. Journal of Fruit Science, 2019, 36(10): 1292–1301. DOI:10.13925/j.cnki.gsxz.2019.10.1292.]
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521 (7553): 436444. DOI:10.1038/nature14539.
- [3] 中国科学院合肥智能机械研究所 智能认知研究组. 葡萄病害图像数据库[DB/OL]. 农业病虫害研究图库. [http://www.icgroupcas.cn/website\\_bchtk/tuku\\_putao.html](http://www.icgroupcas.cn/website_bchtk/tuku_putao.html). [Intelligent Cognition Group, Hefei Institute of Intelligent Machines, Chinese Academy of Sciences. Image Database of Grape Diseases [DB/OL]. Image Database for Agricultural Diseases and Pests Research. [http://www.icgroupcas.cn/website\\_bchtk/tuku\\_putao.html](http://www.icgroupcas.cn/website_bchtk/tuku_putao.html).]

[4] 陈雷, 袁媛. 大田作物病害识别研究图像数据集[J/OL]. 中国科学数据, 2019, 4(4). (2019-06-11). DOI:10.11922/csdata.2019.0008.zh. [CHEN L, YUAN Y. An image dataset for field crop disease identification[J/OL]. China Scientific Data, 2019, 4(4). (2019-06-11). DOI:10.11922/csdata.2019.0008.zh.]

## 论文引用格式

袁媛, 陈雷. IDADP-葡萄病害识别研究图像数据集[J/OL]. 中国科学数据, 2022, 7(1). (2022-02-18). DOI: 10.11922/11-6035.csd.2021.0077.zh.

## 数据引用格式

袁媛, 陈雷. IDADP-葡萄病害识别研究图像数据集[DS/OL]. Science Data Bank, 2021. (2021-11-29). DOI: 10.11922/sciencedb.j00001.00311.

# An image dataset for IDADP-grape disease identification

YUAN Yuan<sup>1,2\*</sup>, CHEN Lei<sup>1,2\*</sup>

1. Institute of Intelligent Machines, HFIPS, Hefei 230031, P.R. China

2. National Basic Science Data Center, Beijing 100190, P.R. China

\*Email: yuanyuan@iim.ac.cn (YUAN Yuan); chenlei@iim.ac.cn (CHEN Lei)

**Abstract:** It is very important to produce an image dataset for machine learning modeling to efficiently identify agricultural diseases and pests. This dataset (about 697 MB) is composed of 3,622 images of 7 varieties about diseases, such as powdery mildew and downy mildew. This dataset is expected to provide valuable basic data resources for the study on grape disease identification and can also serve as a standard library for machine learning modeling in the overall situation of big data. It is of great significance for practical application in promoting the development of agricultural disease identification.

**Keywords:** IDADP; grape disease; image identification; machine learning; training sample

### Dataset Profile

<b>Title</b>	An image dataset for IDADP-grape disease identification
<b>Data corresponding author</b>	YUAN Yuan (yuanyuan@iim.ac.cn), CHEN Lei (chenlei@iim.ac.cn)
<b>Data authors</b>	YUAN Yuan, CHEN Lei
<b>Time range</b>	2018 – 2021
<b>Geographical scope</b>	China
<b>Data volume</b>	697 MB
<b>Data format</b>	*.JPG
<b>Data service system</b>	< <a href="http://www.icgroupcas.cn/website_bchtk/tuku_putao.html">http://www.icgroupcas.cn/website_bchtk/tuku_putao.html</a> > < <a href="http://www.doi.org/10.11922/sciencedb.j00001.00311">http://www.doi.org/10.11922/sciencedb.j00001.00311</a> >



<b>Sources of funding</b>	National Natural Science Foundation of China (31871521, 32071901); “Machine Learning Dataset for Agricultural Image Caption” in National Basic Science Data Center (NBSDC-DB-20).
<b>Dataset composition</b>	The dataset contains 7 varieties of grape disease images, each of which corresponds to a folder, including grapevine powdery mildew, grape mosaic virus disease, rhizopus stolonifer, botrytis cinerea, grape ulcer disease, grape downy mildew, and grape sour rot. Each folder contains the original JPG files named by pipeline number of the disease image and the intro.txt file which introduces the basic information of the disease and its control methods. This dataset is composed of 3,622 images in the format of JPG.