

Scientific Data

<https://doi.org/10.1038/s41597-026-07431-9>

Article in Press

A curated dataset of 3,477 high-resolution Grapevine (*Vitis vinifera*) leaf images for automated detection of Black Rot, Esca, and Leaf Blight diseases

Received: 6 October 2025

Accepted: 9 May 2026

Cite this article as: Gayakwad, M., Mulla, T., Parashar, D. *et al.* A curated dataset of 3,477 high-resolution Grapevine (*Vitis vinifera*) leaf images for automated detection of Black Rot, Esca, and Leaf Blight diseases. *Sci Data* (2026). <https://doi.org/10.1038/s41597-026-07431-9>

Milind Gayakwad, Tahseen Mulla, Deepak Parashar, Anisa Shikalgar, Prashant Chavan, Sachin Wakurdekar, Rohini Jadhav, Amol Kadam, Priyanka Paygude & Rahul Joshi

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A curated dataset of 3,477 high-resolution Grapevine (*Vitis vinifera*) leaf images for automated detection of Black Rot, Esca, and Leaf Blight diseases

Tahseen Mulla¹⁺, Anisa Shikalgar^{2+*}, Deepak Parashar^{3+*}, Milind Gayakwad^{4+*}, Prashant Chavan⁵, Sachin Wakurdekar⁶, Rohini Jadhav⁷, Amol Kadam⁸, Priyanka Paygude⁹, Rahul Joshi¹⁰

^{1, 2}Annasaheb Dange College of Engineering and Technology, Ashta, Shivaji University, India.

³Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India.

^{4,5,6,7,8,9}Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India.

¹⁰Symbiosis Institute of Technology, Pune, Symbiosis International (Deemed University), Pune, India.

*Corresponding author(s). E-mail(s): annu.shikalgar@gmail.com, deepak.parashar@manipal.edu, mdgayakwad@bvucoep.edu.in

Contributing authors: mullatahseen@gmail.com, rbjadhav@bvucoep.edu.in, sbwakurdekar@bvucoep.edu.in, pschavan@bvucoep.edu.in, akkadam@bvucoep.edu.in, pspaygude@bvucoep.edu.in, rahulj@sitpune.edu.in

+These authors contributed equally to this work.

Abstract

We introduce the Grapevine (*Vitis vinifera*) Leaf Image Dataset (GVLiD), a carefully curated set of 3,477 annotated images of Grapevine (*Vitis vinifera*) leaves to catalyze research in computer vision and plant pathology. Whereas the PlantVillage and Hermos datasets, for instance, contain mainly scanned or laboratory-acquired leaves, GVLiD features vineyard in situ images along with detailed metadata (GPS, lighting, weather, and device model) and expert-verified annotations. To measure the reliability of the annotation, label consistency was very high ($\kappa = 0.86\text{--}0.92$; 95% CI) as assessed by inter- and intra-rater agreement. Besides Indian viticulture, the dataset also aims to support the field of foliar disease detection in precision agriculture and ML benchmarking, which face significant challenges due to variable illumination and natural leaf backgrounds under field conditions. GVLiD is intended to enable worldwide, reproducible, real-world testing of AI systems for crop disease monitoring.

Background & Summary

Grapes (*Vitis vinifera*) are among the most economically valuable fruit crops globally, with India being one of the leading countries in the production and export of the fruit [1]. Indian grape-growing areas, especially in Maharashtra and other subtropical regions, are highly vulnerable to fungal diseases such as Black Rot, Esca, and Leaf Blight [2][3]. Detection of these diseases in their initial stages and accurately

is the only way to protect production, minimize chemical usage, and maintain the viability of grape growing [4][5]. Machine learning (ML) and deep learning (DL) methods have been effective for the automatic identification of diseases. However, their performance is highly dependent on the existence of high-quality, diverse, and properly annotated image datasets [6][7]. Several publicly available datasets exist, with PlantVillage being the largest and most popular [8][9], offering more than 50,000 leaf images of various crops. Despite PlantVillage has been a great resource for plant disease recognition research, the limitations of the dataset have also been pointed out: (i) the majority of the pictures are taken against a controlled, uniform background, (ii) most pictures come from temperate areas, and (iii) accompanying data such as GPS, environmental context, and acquisition details are missing [10][11]. These points hinder the applicability of models trained on such data to practical field conditions, especially in tropical viticulture.

Table 1 Comparative Analysis

Dataset	Crop(s)	Size (images)	Geography	Background	Metadata richness	Diseases included
PlantVillage [12]	Multiple crops	>50,000	Temperate, USA-based	Mostly controlled	None	38+ plant diseases
Hermos [13][14]	Grapevine (Vitis vinifera)	2,500	Europe	In-field	Limited	Downy, powdery mildew
Sugarcane Leaf [15][16]	Sugarcane	3,000	India (subtropical)	Mixed	None	Red rot, smut, rust
VegNet [17][18]	Vegetables	2,600	India	Mixed	Limited	15 diseases across 6 crops
GVLiD (ours)	Grapevine (Vitis vinifera)	3,477	India (tropical)	In-situ + ex-situ	Rich (GPS, env., device)	Black Rot, Esca, Leaf Blight

Other projects include Hermos, a European Grapevine (Vitis vinifera) dataset, which concentrates on downy and powdery mildew; the Sugarcane Leaf Dataset from India [19][20], which mainly focuses on monoculture crops; and VegNet, which has images of vegetable diseases with a moderate variety of classes but scarce metadata. In contrast to these materials, GVLiD stands out by integrating a dual acquisition strategy (in situ vineyard conditions and ex situ controlled backgrounds) [21][22]. Locating GVLiD in this worldwide context reveals its value. Whereas datasets such as PlantVillage have provided indispensable baseline data for concept validation, GVLiD is a timely solution to the shortage of geographically diverse, metadata-rich, and field-validated datasets [23][24][25]. In addition to ensuring local applicability for Indian grape farming, it serves as a test environment to investigate domain adaptation, transfer learning, and machine learning model robustness across different climate zones [27][28]. As viticulture is being established in tropical and subtropical regions globally, GVLiD offers worldwide-significant insights and resources that support precision agriculture, sustainable practices, and cross-regional disease surveillance.

Methods

1. Data Collection

Data were collected from 34 vineyards in Tasgaon, Sangli district, Maharashtra, India (July-November 2023). A dual acquisition strategy was followed: For each leaf, photos of both the adaxial (upper) and abaxial (lower) sides were taken. All disease labels were prepared by a highly experienced agricultural expert with more than 18 years of viticulture experience. Data acquisition was done according to a documented protocol to ensure repeatability.

1.1 Capture specifications

- Device models: Samsung Galaxy S21 Ultra and Redmi Note 10 5G
- Capture angle: 20–45° relative to leaf plane
- Distance: 25–40 cm
- Lighting: sunny, overcast, shaded, dusk, and controlled LED
- Background: natural vineyard (in situ) and neutral grey board (ex situ)

Both adaxial and abaxial leaf surfaces were photographed.

The creation of specialized image datasets is an indispensable part of developing machine learning applications in agriculture, especially in areas such as precision farming and automated disease detection [16]. This dataset was created to be a top-quality resource for these purposes. The timeline showing data curation and preprocessing is in Figure 1.

1.2 Data Collection Protocol

An agricultural expert with over 18 years of experience in viticulture verified the category of each leaf sample before imaging, ensuring label accuracy from the point of origin. A standardized acquisition protocol was followed to ensure reproducibility.

Figure 1. Image Acquisition and Data Organization

2. **Data Curation** – Data is curated during the Project Planning & Site Selection (July 2023), Image Acquisition (August 2023 - September 2023), and Data Curation (October 2023) phases. Details are mentioned below-

Images were stored in JPEG format with a compression quality level of Q=95, which was experimentally determined to give a good trade-off between storage efficiency and visual fidelity. For image acquisition, a dual-lighting strategy was used. In situ images were taken in the vineyard under different lighting conditions, including sunny, overcast, shaded, and dusk. Using this approach, the dataset obtained real-world variability such as background clutter and uneven illumination. Besides, ex situ images were taken in a controlled setting with a neutral gray background, and the illumination was provided by LED lights to minimize the appearance of shadows and reflections. Overall, this dual approach greatly increases generalizability: on the one hand, in situ data facilitates the development of models that perform well in the field; on the other hand, ex situ images serve as clean diagnostic references for training and benchmarking.

2.1 Image Annotation:

1. All images were labeled by an agricultural expert (Name: Mr. Avadhut Disale, Mr. Omkar Bhosale; Affiliation: Agriculture Consultant; 18 years' experience in viticulture). Identify primary lesion type (necrotic spots, rotting margins, melanose).
2. Check lesion distribution (interveinal vs. marginal).
3. Assess presence of secondary symptoms (discoloration, pustules).
4. Cross-check with expert notes recorded at sampling.

Primary annotation was done by only one expert focusing on the domain as the main rationale; the reliability was gauged by inter- and intra-rater validations on a data subset. Although full dual annotation was not carried out, the very good levels of inter- and intra-rater agreement indicate that label consistency is reliable across the entire dataset. All the images were field-annotated by an agri-expert. Labeling was carried out via a decision tree (Figure) structure that dictated the procedure by first recognizing the main lesion type (necrotic spots, rotting margins, melanose), then analyzing lesion location (interveinal versus marginal), looking at secondary symptoms such as chlorosis, pustules, or blight patterns, and lastly, comparing with expert notes that were made at the time of sampling. This consistent annotation structure not only reduces subjectivity but also makes it easy to maintain transparency if the dataset is reused later.

2 Data Curation:

The raw dataset initially contained 3792 images. Following a three-stage curation pipeline:

1. Integrity check: 15 corrupt files removed (<0.4%).
2. Quality check: 200 blurred/overexposed images excluded (5.3%).
3. Duplicate removal: 100 near duplicates identified by perceptual hashing (2.6%).

After curation, the final release contained 3,477 high-quality images, representing 91.7% of the original collection.

3 Data Preprocessing (November 2023)

Image Resizing: All images are standardized to a uniform resolution of 1080 x 1080 pixels to ensure uniformity for further processing. The data distribution is shown in the figure. 2.

Data Organization - Final Dataset

A curated, annotated, and instantly usable resource with 3,477 photos. Three different versioning stages. Originally, version 1 was the basic dataset, version 2 included the metadata, version 3 included the metadata with the details of image capture in metadata, version 4 covers the ISO standards of time of

capturing the images and completeness of data, version 5 covers the image name, and sensor to capture the images of Grapevine (*Vitis vinifera*).

Figure. 2. Data distribution

The data set shows a fair amount of class imbalance (the healthy one is overrepresented by approximately 1.65 times). As the number of healthy images increases, considering all classes in a weighted manner helps mitigate class imbalance.

Data Records

Data records represent a logical, programmatically reachable layout, like the data structure in Figure 4. Data organization: The Grape Leaf Dataset contains high-resolution images of grape leaves. Every image in this dataset is a JPG file with 1080×1080 pixels. There are four classes in the dataset: three disease categories and one healthy-leaf category (Figure 1). In addition to healthy leaves, the database also lists images of various Grapevine (*Vitis vinifera*) diseases, such as Black Rot, Esca, and Leaf Blight (Table 1). To facilitate quick access and use of the visual content, there are separate folders for each disease category. Various disease symptoms, along with natural variations across different grapevine-growing sites, were collected during field visits by experts. The directory structure used for uploading to the Mendely platform is depicted in Figure 4. The dataset, metadata, and other files needed for data reproduction are available, and the validation script is on GitHub.

Figure 3. Sample Images of different Grapevine (*Vitis vinifera*) leaves (Unhealthy, Healthy)

Figure 3 presents some content(samples) of image(s) depicting various symptoms of the disease(s). Visualizing such features can be a useful guide to understanding how the disease works, both visually and through small-scale, first-hand reproductions of environments. Table 2 lists the location, surrounding environment, and camera type used to take the picture(s). This kind of information is a little glimpse into ex-situ (outside the natural location) image capturing.

The dataset is organized into three main components: **(1) Image data, (2) Metadata, and (3) Documentation/Validation files**

Root Directory: GVLiD/

This is the main dataset container that ensures all components required for reproducibility are available in one place.

1. Image Data The GVLiD dataset structures its image data by categorizing the images of Grapevine leaves into different class-wise folders, like `black_rot`, `esca`, `healthy`, and `leaf_blight`. Each folder contains Grapevine (*Vitis vinifera*) leaf images at 1080×1080 pixels. This level of directory categorization directly corresponds to disease labels, ensuring that every image is clearly linked to a single class. Maintaining separate folders for each disease category reduces the likelihood of label errors and also helps with data

loading, preprocessing, and model training. So, this dataset is very useful for tasks such as classification, benchmarking, and comparative analysis of plant disease detection.

Figure 4. Directory Structure of the Grapevine (*Vitis vinifera*) Leaf Dataset

1. Metadata

Metadata is created during the moment of taking the picture: information that does not change, like device details, is already saved, while information that changes, like exact location, time, and type of image, is input during capture. Every photo has a metadata record in JSON format. Metadata fields include vineyard ID, GPS coordinates, capture time, lighting conditions, weather conditions, device specifications, plant health status, and disease type. This approach improves reproducibility and allows for the exploration of the relationship between the environment and disease manifestation. The exact data can be found in Metadata. JSON is referred to immediately after Table 2. Metadata is written in JSON format to document parameters such as lighting conditions, plant health, and the device name.

Table 2. Metadata sample

Metadata.json	Explanation
<pre>{ "Image_ID": "IMG_0001", "image_filename": "black rot001.jpg", "Location": { "Vineyard": "Vineyard_22", "GPS": { "Latitude": 18.785238, "Longitude": 73.994949 } }, "Environmental_Conditions": { "Time_of_Capture": "2023-06-15T15:13:00+05:30", "Lighting": "Overcast", "Weather": "Humid" }, "Plant_Health_Status": "Diseased", "Disease_Type": "Black Rot", "Device_Specifications": { "Device_Name": "Redmi Note 10 5G", "Sensor": "Sony IMX 582 1/2rr", "OS": "Android" } }</pre>	Image_ID (integer) — serial number indicates the image number image_filename (string) – Name of the image indicates the name of the disease and a unique number. Location Vineyard (string) — vineyard identifier (e.g., Vineyard_22) GPS. Latitude (float) — decimal degrees (WGS84) Longitude (float) -decimal degrees Time_of_Capture (ISO 8601 string) — e.g., 2023-08-15T14:35:00+05:30 Lighting (controlled vocabulary: sunny, overcast, shade, dawn/dusk, artificial) Weather (string) — short descriptor (e.g., humid) Plant_Health_Status (enum: Diseased, Healthy) Disease_Type (enum: Black Rot, Esca, Leaf Blight, Healthy) Device_Specifications Device_Name (string) — e.g., Redmi Note 10 5G Sensor (string) — e.g., Sony IMX582 OS (string) – e.g. Android

This is the actual sample of the metadata available at the dataset organization.

Technical Validation

The dataset went through a multi-step validation process.

Expert verification - The plot agronomist, physically present at the site, verified all the labels with us again, and they are 100% accurate. The image labeling was accurate, as confirmed by the feedback of the Agricultural Experts.

Balanced classes - Distributing classes evenly reduces the risk of model bias.

Image standardization: Every picture was resized to the same size beforehand. The pictures were processed to reveal the features that are needed to distinguish the disease. We have subjected GVLiD to a multi-layer inspection, including checks of dataset integrity, metadata completeness, image quality, and annotation reliability. All these inspections show that the dataset can be safely applied in machine learning, computer vision, and plant pathology research.

Image Integrity

SHA-256 checksums for each image are provided in docs/checksums.txt, allowing users to independently verify file integrity.

Image Quality Assessment

To ensure consistency and suitability for automated analysis, we computed basic image quality metrics:

1. **Resolution:** All images standardized to 1080×1080 pixels.
2. **Compression:** JPEG, Q=95 (empirically optimized to balance fidelity and size).
3. **Brightness:** Mean luminance = 127.3 (SD = 18.5).
4. **Sharpness:** Mean Tenengrad score = 8.7×10^3 .
5. **Blurriness:** 73 images (2.1%) flagged as potentially blurry by automated heuristics; manual review confirmed that most retained diagnostic value.

Metadata Completeness

The metadata file (metadata.json) provides rich contextual information. Completeness was quantified across 3,477 entries:

Table 3. Completeness of the Metadata

Sr_No	Field Name	Available / Total	Completeness (%)
1	Sr_No	3477	100.0
2	Plant_Health_Status	3477/3477	100.0
3	Disease_Type	3477/3477	100.0
4	Location.Vineyard	3477/3477	100.0

5	Location.GPS.Latitude	3477/3477	100.0
6	Location.GPS.Longitude	3477/3477	100.0
7	Environmental_Conditions.Time_of_Capture	3477/3477	100.0
8	Environmental_Conditions.Lighting	3477/3477	100.0
9	Environmental_Conditions.Weather	3477/3477	100.0
10	Device_Specifications.Device_Name	3477/3477	100.0
11	Device_Specifications.Sensor	3477/3477	100.0
12	Device_Specifications.OS	3477/3477	100.0

As mentioned in Table 3, the complete statistics are summarized in docs/metadata_completeness.csv.

Annotation Reliability

We assessed label reliability through inter- and intra-rater agreement:

1. A random subset of 200 images was re-annotated after four weeks by the same expert (intra-rater) and by a second expert (inter-rater).
2. **Inter-rater agreement:** Cohen's $\kappa = 0.86$, 95% CI [0.82–0.90], indicating *substantial agreement*.
3. **Intra-rater agreement:** Cohen's $\kappa = 0.92$, 95% CI [0.89–0.95], indicating almost perfect agreement.
4. Disagreements (<5% of subset) were resolved by consensus, and approximately 3% of images in the dataset are flagged as “uncertain” in metadata.

Class Balance

To identify potential biases, we quantified the distribution of classes:

1. **Black Rot:** 809 images
2. **Esca:** 888 images
3. **Leaf Blight:** 672 images
4. **Healthy:** 1109 images

While the dataset is relatively balanced, there is a slight skew toward Healthy Images (1109), which should be considered when using the dataset. Stratified sampling is recommended, with 70% of the data for training, 15% for validation, and 15% for testing. Stratified sampling is recommended.

Reproducibility

All validation scripts (image-quality metrics, metadata-completeness checks, and annotation-reliability calculations) are publicly available in the GVLiD GitHub repository. This ensures full reproducibility of all validation results reported here.

```
git clone https://github.com/MilindGayakwad/DNN
cd DNN/GVLiD
python validate_metadata.py
python compute_quality_metrics.py
```

This script generates a validation report including completeness statistics, missing values, and schema compliance checks.

```
python validate_metadata.py --input metadata.json --output report.json
```

Limitations: GVLiD dataset was acquired in Tasgaon (Sangli district) of Maharashtra and hence primarily reflects the disease symptoms and the grapevine varieties of that region and season. Therefore, if a model is trained only on GVLiD data, it is likely to perform worse on leaves from other climate zones or cultivars. GVLiD has included both in situ and ex situ photos to address domain variability to some extent; however, further geographic expansion is required to achieve worldwide generalization. Despite GVLiD offering a varied collection of real-life grapevine (*Vitis vinifera*) leaf images, some drawbacks still exist. For one, the dataset was gathered from a single geographical location (Tasgaon, Sangli district, Maharashtra, India), which might limit the representation of grapevine (*Vitis vinifera*) cultivars and viticultural environments outside this region. Two, the data were collected for one single growing season (July-November 2023), excluding seasonal variation in disease expression. Three, the pictures were taken using two different smartphone camera sensors, which could lead to slight differences in color tone and sensor properties. Though this discrepancy enhances the machine learning models' resilience, it could also lead to device-specific biases. Finally, the dataset includes multiple disease stages but does not explicitly label the progression stage of each infection, which may be relevant for studies focusing on disease severity estimation.

Usage Note

GVLiD is designed to facilitate ML/DL research on Grapevine (*Vitis vinifera*) disease detection. Its organized directory structure makes it directly compatible with frameworks such as TensorFlow and PyTorch. Researchers can explore tasks such as image classification, transfer learning, and domain adaptation. The metadata allows additional studies on environmental factors, generalization, and cross-regional disease analysis. The dataset also provides a foundation for smartphone-based viticulture applications.

Data Availability

The dataset is available at Mendeley Data [29]

Code Availability

The link to the GitHub project is <https://github.com/MilindGayakwad/DNN>

Author Contributions

Conceptualization: T. Mulla, A. Shikalgar.

Supervision: D. Parashar, M. Gayakwad.

Data Curation: R. Jadhav, A. Kadam, P. Chavan.

Methodology: S. Wakurdekar, P. Paygude.

Writing – Original Draft: R. Joshi.

Writing – Review & Editing: All authors

Validation: Agricultural expert (15+ years experience)

Competing Interests

The authors declare no competing interests.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution-Non-commercial-No Derivatives 4.0 International License, which permits any non-commercial use, sharing, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third-party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>. **Reprints and permissions**

Ethics Statement

Not applicable (no human or animal data).

REFERENCES

- [1] Suryawanshi, Y., Patil, K. & Chumchu, P. VegNet: dataset of vegetable quality images for machine learning applications. Data in Brief 45, 108657 (2022). <https://doi.org/10.1016/j.dib.2022.108657>
- [2] Kaur, S., Pandey, S. & Goel, S. Plant disease identification and classification through leaf images: a survey. Archives of Computational Methods in Engineering 26, 507–530 (2019).
- [3] Hemalatha, N.K. et al. Grapes leaf disease detection through deep learning. In: Deep Learning for Sustainable Agriculture, 297–323 (Academic Press, 2022).

- [4] Thite, S., Suryawanshi, Y., Patil, K. & Chumchu, P. Sugarcane leaf dataset: a dataset for disease detection and classification for machine learning applications. *Data in Brief* (2024). <https://doi.org/10.1016/j.dib.2024.110268>
- [5] Suryawanshi, Y. et al. Yoga dataset: a resource for computer vision-based analysis of yoga asanas. *Data in Brief* 48, 109257 (2023). <https://doi.org/10.1016/j.dib.2023.109257>
- [6] Azlah, M.A.F. et al. Review on techniques for plant leaf classification and recognition. *Computers* 8(4), 77 (2019). <https://doi.org/10.3390/computers8040077>
- [7] Shorten, C. & Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *Computers and Electronics in Agriculture* 147, 70–90 (2018). <https://doi.org/10.1016/j.compag.2017.12.024>
- [8] Radhika, K. & Vinayakumar, R. DeepLeaf: an optimized deep learning approach for automated recognition of Grapevine (*Vitis vinifera*) leaf diseases. *Neural Computing and Applications* (2025). <https://doi.org/10.1007/s00521-025-11038-3>
- [9] Mansour, R.F. et al. Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. *Journal of Ambient Intelligence and Humanized Computing* (2024). <https://doi.org/10.1007/s41348-024-00896-z>
- [10] Kunduracıoğlu, I. & Paçal, İ. Advances in grape leaf disease detection using deep learning. *Journal of Plant Diseases and Protection* 131(1), 45–60 (2024). <https://doi.org/10.1007/s41348-024-00896-z>
- [11] Zhuang, L. et al. Grape leaf disease identification using improved deep convolutional neural networks. *Journal of Ambient Intelligence and Humanized Computing* 11, 1–10 (2020). <https://doi.org/10.1007/s12652-020-02104-2>
- [12] Li, W. et al. Deep learning-based system for early symptom recognition of Grapevine (*Vitis vinifera*) diseases. *Journal of Imaging* 9(3), 63 (2023). <https://doi.org/10.3390/jimaging9030063>
- [13] Wang, Y. et al. Deep learning based automatic grape downy mildew detection. *Frontiers in Plant Science* 13, 872107 (2022). <https://doi.org/10.3389/fpls.2022.872107>
- [14] Javidan, S.M., Banakar, A., Asefpour Vakilian, K. & Ampatzidis, Y. Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. *Smart Agricultural Technology*, 3, 100081 (2023). <https://doi.org/10.1016/j.atech.2022.100081>
- [15] Molina, D.C. et al. Assessment of downy mildew in Grapevine (*Vitis vinifera*) using computer vision. *OENO One* 56(4), 1–10 (2022). <https://doi.org/10.20870/oeno-one.2022.56.4.5359>
- [16] García-Martínez, H. et al. Hermos: an annotated image dataset for grape leaf disease detection.

Journal of Information Science 48(1), 123–135 (2022). <https://doi.org/10.1177/01655515221091892>

[17] Zhang, H. et al. Early detection of grape downy mildew using YOLOv3. Computers and Electronics in Agriculture 186, 106169 (2021). <https://doi.org/10.1016/j.compag.2021.106169>

[18] Zhang, D. et al. Lightweight network for grape disease detection on mobile devices. Computers and Electronics in Agriculture 205, 107639 (2023). <https://doi.org/10.1016/j.compag.2023.107639>

[19] Liu, J. et al. Grape leaf disease detection using ensemble CNN models. Scientific Reports 12, 4762 (2022). <https://doi.org/10.1038/s41598-022-08762-3>

[20] Zhao, Y. et al. Machine learning techniques in grape disease detection using thermal imaging. Remote Sensing 12(15), 2453 (2020). <https://doi.org/10.3390/rs12152453>

[21] Shi, Y. et al. Image-based recognition of grape leaf disease using MobileNetV2. Sensors 21(17), 5817 (2021). <https://doi.org/10.3390/s21175817>

[22] He, Z. et al. Grapevine (*Vitis vinifera*) disease recognition using attention-based CNN. Computers and Electronics in Agriculture 175, 105543 (2020). <https://doi.org/10.1016/j.compag.2020.105543>

[23] Singh, V. et al. Grape leaf disease classification using transfer learning. Multimedia Tools and Applications 80, 11911–11931 (2021). <https://doi.org/10.1007/s11042-020-10135-0>

[24] Phadikar, S. et al. Grape leaf disease detection using deep neural networks. AI Open 3, 107–117 (2022). <https://doi.org/10.1016/j.aiopen.2022.01.002>

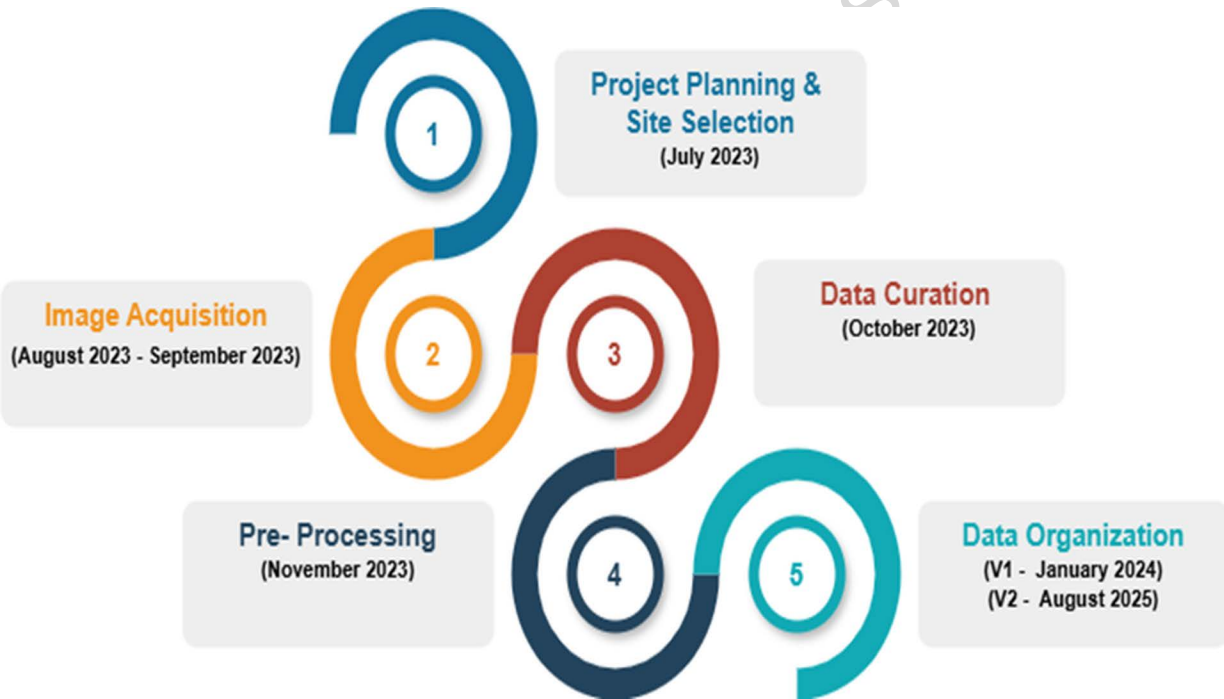
[25] Mahmoud, S. et al. Vision-based monitoring for Grapevine (*Vitis vinifera*) powdery mildew using YOLOv5. Computers in Biology and Medicine 156, 106666 (2023). <https://doi.org/10.1016/j.compbiomed.2023.106666>

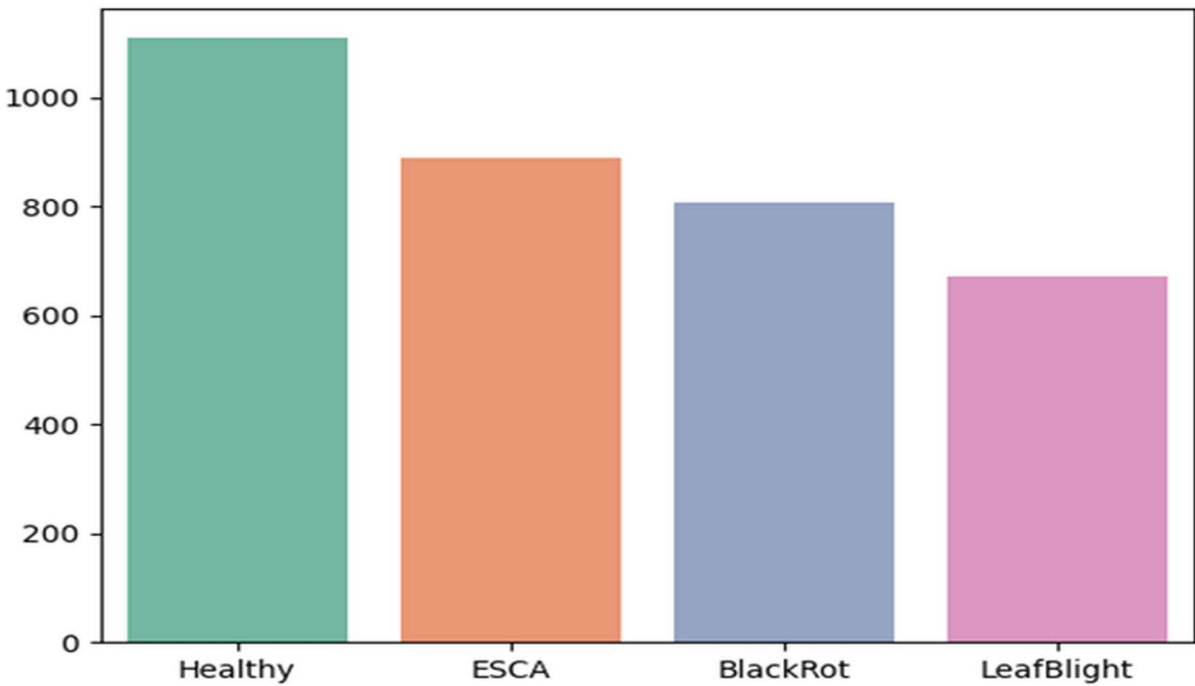
[26] Zhang, L. et al. Multiscale fusion-based deep network for grape disease detection. Pattern Recognition Letters 174, 27–34 (2024). <https://doi.org/10.1016/j.patrec.2023.12.005>

[27] Habib, R. et al. CNN-RNN hybrid framework for Grapevine (*Vitis vinifera*) disease detection. Neural Computing and Applications 33, 7985–7998 (2021). <https://doi.org/10.1007/s00521-020-05235-3>

[28] Mansour, R.F. et al. Advances in deep learning for grape disease classification. Journal of Ambient Intelligence and Humanized Computing 15(3), 123–135 (2024). <https://doi.org/10.1007/s41348-024-00896-z>

[29] Shikalgar, Anisa; Savalkar, Ayush; Bhasme, Avishkar; Chavan, Snehal; Nikam, Vaishnavi (2026), DOI:10.17632/wkymf8bhcg.5







GVLiD/

└─ black_rot/

└─ esca/

└─ leaf_blight/

└─ healthy/

└─ METADATA/

└─ └─ metadata.csv

└─ docs/

└─ └─ checksums.csv

└─ └─ image_quality_metrics.csv

└─ └─ metadata_completeness.csv

└─ └─ validation_report.json