Snap and Diagnose: An Advanced Multimodal Retrieval System for Identifying Plant Diseases in the Wild

Tianqi Wei, Zhi Chen, Xin Yu
The University of Queensland
Brisbane, Australia
{tianqi.wei,zhi.chen,xin.yu}@uq.edu.au

Abstract

Plant disease recognition is a critical task that ensures crop health and mitigates the damage caused by diseases. A handy tool that enables farmers to receive a diagnosis based on query pictures or the text description of suspicious plants is in high demand for initiating treatment before potential diseases spread further. In this paper, we develop a multimodal plant disease image retrieval system to support disease search based on either image or text prompts. Specifically, we utilize the largest in-the-wild plant disease dataset PlantWild, which includes over 18,000 images across 89 categories, to provide a comprehensive view of potential diseases relating to the query. Furthermore, cross-modal retrieval is achieved in the developed system, facilitated by a novel CLIP-based vision-language model that encodes both disease descriptions and disease images into the same latent space. Built on top of the retriever, our retrieval system allows users to upload either plant disease images or disease descriptions to retrieve the corresponding images with similar characteristics from the disease dataset to suggest candidate diseases for end users' consideration.

CCS Concepts

• **Computing methodologies** → **Multimodal Retrieval**.

Keywords

Plant disease recognition, Multimodal image retrieval, Vision language models

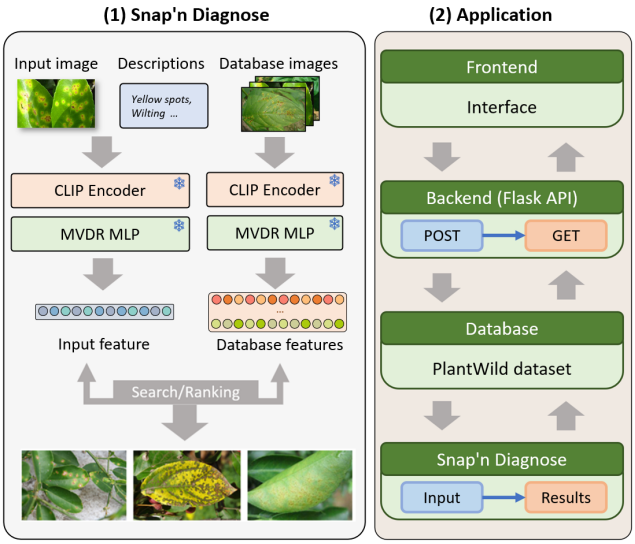

Figure 1: Overview of Snap'n Diagnose. We leverages CLIP [8] and MVPDR [11] to extract visual/text features from images and texts and conduct multimodal image retrieval for identifying plant disease in the wild.

1 Introduction

With the global population on the rise, the demand for food continues to escalate [5]. Plant diseases significantly reduce crop yield reduction, inflicting economic losses exceeding $200 billion annually [1]. Plant disease recognition plays a vital role in crop protection against diseases, as it directly impacts agricultural sustainability and global food security. Traditionally, plant disease recognition relies on experienced farmers or agricultural experts for manual identification, but these practices are time-consuming, costly, and not always available. In this context, automatic disease recognition with machine learning approaches has drawn much attention in the plant pathology community. While existing methods have achieved promising results on in-laboratory images [3, 6, 9], their performance significantly declines when applied to images captured in the wild. Furthermore, farmers often need to match observed symptoms, such as "yellow spots on leaves" or "wilting flowers", with corresponding images. It is thus desirable to achieve cross-modal disease retrieval with textual queries.

To facilitate this need, current plant disease image retrieval systems predominantly support unimodal (image-only) queries [2, 10] and are limited by plant types [2, 12] and inability to handle in-the-wild images [3, 6, 7, 9]. These constraints hinder their practical utility in diverse agricultural settings.

To address these limitations, in this paper, we propose a multimodal image retrieval system for plant diseases in the wild. Specifically, to accommodate the needs for diverse plant diseases, we construct our retrieval database with the world-largest plant disease dataset PlantWild [11]. It includes over 18,000 in-the-wild plant images across 89 classes. Notably, PlantWild dataset provides diverse textual descriptions for each disease type. Further, to facilitate cross-modal queries, we develop a CLIP-based [8] retrieval method that projects both images and textual descriptions of each disease into the same latent space. Such a shared latent space allows users to retrieve the image samples closest to the query. Built upon this retrieval model, we then develop a retrieval system **Snap'n Diagnose** to provide an interface for retrieval interaction.

Snap'n Diagnose offers a user-friendly interface that simplifies the retrieval process. Users can either type in a textual description of observed symptoms or upload a photo of the affected plant. The system then transforms this input into query representations and calculates similarities with images in the database. The results,

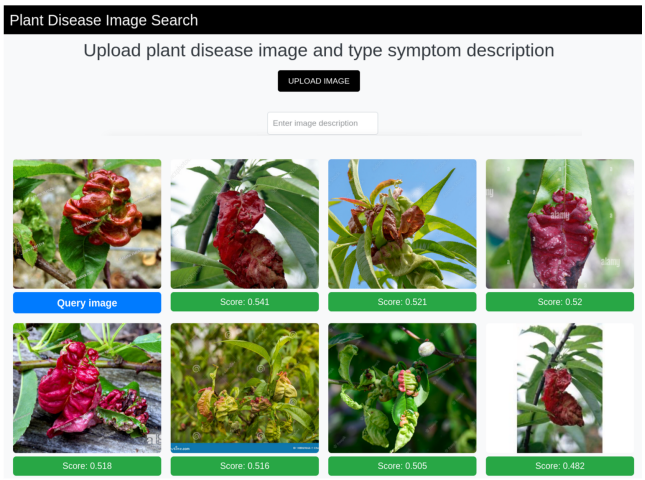

Figure 2: A screenshot of our system interface. After receiving the image/textual queries, Snap'n Diagnose results will be returned and ranked in order of cosine similarity.

ranked by relevance, are promptly displayed, providing users with reliable and actionable insights into potential plant diseases.

2 System

2.1 Methodology

The retrieval model in Snap'n Diagnose is based on the MVPDR method [11], which is designed for in-the-wild plant disease recognition and has proved effective in adapting images across different environments. Therefore, it is suitable for searching plant disease images. Besides, it also benefits from the image search engine [4].

Snap'n Diagnose extracts and stores visual/text features with CLIP encoders and MVPDR's fine-tuned MLP. These obtained image features preserve a wide range of plant disease features due to the pre-trained CLIP and the fine-tuned MLP. During the inference process, a given input image or textual query will first be extracted into a feature vector. Afterward, the cosine similarity between the vector and all the stored features can be acquired. The outcome is a ranked list that reflects the most pertinent results corresponding to the input image. The workflow is presented in Figure 1.

2.2 Implementation

The retrieval system has been designed for users to interactively conduct multimodal image retrieval, as presented in Figure 2. The system consists of a backend component to perform multimodal retrieval and a frontend interface for user interaction.

Backend. We deploy Snap'n Diagnose with a Flask API in the backend. The API receives the uploaded images/texts as queries and performs inference, searching for images with similar symptoms in a database for responses. We leverage PlantWild [11] as our database, which is the largest plant disease image dataset that includes a vast collection of images captured in diverse environments. Specifically, we extract visual features using CLIP's image encoder and MVPDR's pre-trained MLP, utilizing these features for inference. This operation not only reduces CPU usage but also removes the time for models to process images. Therefore, our system has a fast response time, providing better user experiences.

Table 1: Performance comparison of multimodal image retrieval for plant diseases between different models.

Methods	Top-1	Top-5	Top-10	mAP
Zero-shot CLIP [8]	40.92	65.75	74.81	68.72
Snap'n Diagnose (ours)	**67.32**	**80.65**	**88.11**	**79.34**

Frontend interface. The interface of our system is presented in Figure 2. It is accessible through web browsers on both PC and mobile devices. The design is straightforward, allowing users to easily use it without extensive knowledge. When encountering unknown plant diseases in the wild, users can simply take a photo or typing the symptom descriptions and upload it via the interface. After receiving the request, the backend will perform cross-modal retrieval to obtain similar image results and then return them to the frontend interface. The result images will be arranged following the query image in descending order of their similarity scores, and the corresponding scores will be displayed below each image.

3 Experiment

We evaluate the performance of the Snap'n Diagnose in plant disease image retrieval and compare its performance with the pre-trained CLIP vision-language model. Experiments are conducted on the PlantWild dataset. According to the results presented in Table 1, our method exhibits excellent performance and consistently outperforms Zero-shot CLIP across all evaluation metrics, including Top-1, Top-5, Top-10 accuracy, and mean Average Precision (mAP). These results underscore the effectiveness of Snap'n Diagnose in retrieving relevant plant disease images, demonstrating that it can offer a practical tool in plant disease recognition.

4 Conclusion

In this paper, we present Snap'n Diagnose, a multimodal image retrieval system designed for identifying plant disease in-the-wild. It addresses the limitations of existing retrieval systems that only support unimodal, laboratory and single plant types. Further, Snap'n Diagnose enables farmers to receive plant diagnosis based on query pictures or the textual description of suspicious symptoms.

References

[1] George N Agrios. 2005. *Plant pathology.*
[2] Douglas Baquero, Juan Molina, Rodrigo Gil, Carlos Bojacá, Hugo Franco, and Francisco Gómez. 2014. An image retrieval system for tomato disease assessment. In *2014 XIX Symposium on Image, Signal Processing and Artificial Vision.* IEEE, 1–5.
[3] David Hughes, Marcel Salathé, et al. 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060* (2015).
[4] Yusuke Matsui, Takuma Yamaguchi, and Zheng Wang. 2020. CVPR2020 Tutorial on Image Retrieval in the Wild. https://matsui528.github.io/cvpr2020_tutorial_retrieval/.
[5] E-C Oerke, H-W Dehne, Fritz Schönbeck, and Adolf Weber. 2012. *Crop production and crop protection: estimated losses in major food and cash crops.* Elsevier.
[6] David Opeoluwa Oyewola, Emmanuel Gbenga Dada, Sanjay Misra, and Robertas Damaševičius. 2021. Detecting cassava mosaic disease using a deep residual convolutional neural network with distinct block processing. *PeerJ Computer Science* (2021), e352.

[7] Yingshu Peng and Yi Wang. 2022. Leaf disease image retrieval with object detection and deep metric learning. *Frontiers in Plant Science* 13 (2022), 963302.

[8] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. 8748–8763.

[9] Muhammad Sharif, Muhammad Attique Khan, Zahid Iqbal, Muhammad Faisal Azam, M Ikram Ullah Lali, and Muhammad Younus Javed. 2018. Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. *Computers and electronics in agriculture* (2018), 220–234.

[10] Gulbir Singh and Kuldeep Kumar Yogi. 2020. A review on recognition of plant disease using intelligent image retrieval techniques. *Asian Journal of Biological Life Science* 9, 3 (2020), 274–285.

[11] Tianqi Wei, Zhi Chen, Zi Huang, and Xin Yu. 2024. Benchmarking In-the-Wild Multimodal Plant Disease Recognition and A Versatile Baseline. In *ACM International Conference of Multimedia*.

[12] Wang Zhijun, Liu Yuefeng, Jiang Meng, Cheng Shuhan, and Wang Yucun. 2015. Research on Image Retrieval of Fruit Tree Plant-Diseases and Pests Based on Nprod. *Intelligent Automation & Soft Computing* 21, 3 (2015), 371–381.