The Journal of Open Source Software

Pynomic: Data extraction library for orthomosaic images in plant breeding trials

Juan Manuel Fiore [1¶] **and Cecilia Bruno** [1,2]

1 Universidad Nacional de Córdoba (UNC) **2** Cátedra de Estadistica y Biometría. ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 31 August 2025
Published: unpublished

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

DRAFT

Summary

Plant breeding is the discipline that develops cultivars that meet certain conditions with the purpose to optimize the resource gathering from plants to society. These conditions can be yield, high protein, resilience to biotic or abiotic stress, fiber, oil, etc. For that, first, genetic variability is created by crossing different individuals. Then their progeny is tested in the field. This part of the process is the most expensive since normally they are tested in multi-environmental trials with multiple repetitions. This generates data to select the best individuals.

One of the crucial steps in plant breeding is the collection of data from field trials to characterize and understand the behavior of each experimental plant variety and choose the best ones that adapt to the desired environments. This task is labor-intensive and thus has high costs. To avoid this situation, the use of drones or unmanned aerial vehicles has expanded due to their capabilities to phenotype thousands of plots in a matter of minutes, relatively low costs, and objective measurements (Shakoor et al., 2017). This opens a field of possibilities to obtain accurate and objective data used to train models for yield predictions (Yuan et al., 2024), length of cycle (Volpato et al., 2021), lodging (Koh et al., 2021), monitoring plant stress, and many other cases (Shakoor et al., 2019). Pynomic automates the data extraction of images from plant breeding fields taken by Unmanned Aerial Vehicles (UAVs).

Statment of need

The raw data generated by UAVs has to be processed to extract meaningful information. Many methods have been developed, but they require manual action, which can make the process slow and tedious. As well, the volume of data handled makes the process resource-demanding.

Pynomic is a Python-based, open-source library that automates data workflows for plant breeding trials. It handles the extraction, analysis, and visualization of data from time-series orthomosaics, enabling researchers to monitor the growth, senescence, and stress patterns of each plot while addressing common challenges in data management. To achieve this efficiency, Pynomic relies heavily on the zarr library(Zarr Development Team, 2024), which provides fast and easy background access to the image matrices.

Pynomic was designed as a versatile data extraction tool. It can generate numerical features, such as Vegetation Indices (VIs) and textural information, using both pre-built and custom user-defined functions. Additionally, it prepares data for image-based models by automating the cropping, orientation, and saving of individual plots as .tiff and .png files.

key features of `Pynomic`

1. Automated Time-Series Phenotyping: Pynomic automates the extraction of plot-level

data from time-series UAV orthomosaics (both RGB and Multispectral), streamlining high-throughput phenotyping workflows.

2. Optimized Computational Efficiency: Significantly reduces RAM usage and disk space requirements for large datasets through its Zarr-based backend and lazy loading capabilities, leading to faster processing times.
3. Flexible Feature Engineering: Offers built-in calculation of common Vegetation Indices (VIs) and Textural Features (TFs), alongside a framework for users to easily define and apply custom feature extraction functions.
4. Integrated Analytical Tools: Includes functionalities for practical applications, such as senescence and maturity prediction models, and facilitates seamless integration with the broader Python machine learning ecosystem.
5. Open-Source and Quality-Focused: Provides a user-friendly, well-documented, and rigorously tested open-source library, developed to high standards to foster reproducibility and accessibility in plant phenotyping research.

Documentation and examples of use has been created.

Pynomic Scheme

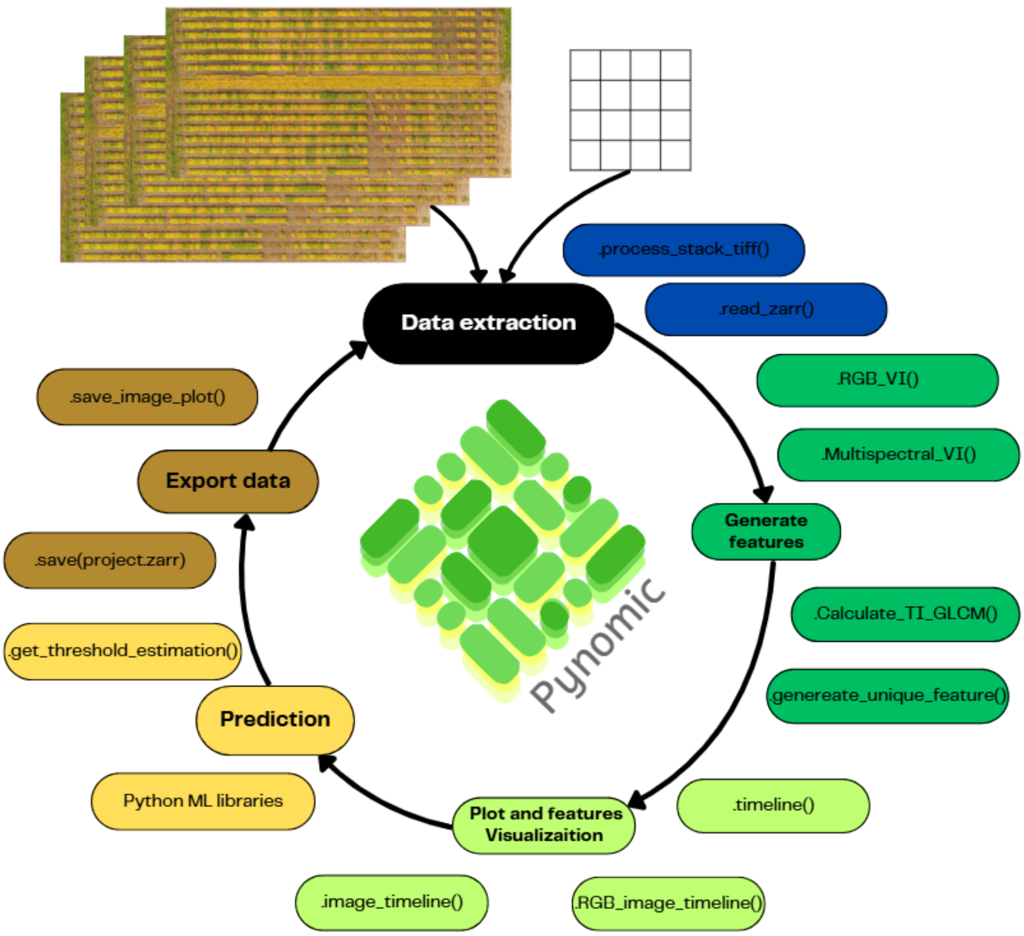

Figure 1: Pynomic Workflow

Fig 1: Display of Pynomic workloflow and functionalities.

References

Koh, J. C. O., Spangenberg, G., & Kant, S. (2021). Automated machine learning for high-throughput image-based plant phenotyping. *Remote Sensing*, *13*(5). https://doi.org/10.3390/rs13050858

Shakoor, N., Lee, S., & Mockler, T. C. (2017). High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Current Opinion in Plant Biology*, *38*, 184–192. https://doi.org/https://doi.org/10.1016/j.pbi.2017.05.006

Shakoor, N., Northrup, D., Murray, S., & Mockler, T. C. (2019). Big data driven agriculture: Big data analytics in plant breeding, genomics, and the use of remote sensing technologies to advance crop productivity. *The Plant Phenome Journal*, *2*(1), 180009. https://doi.org/https://doi.org/10.2135/tppj2018.12.0009

Volpato, L., Dobbels, A., Borem, A., & Lorenz, A. J. (2021). Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding. *The Plant Phenome Journal*, *4*(1), e20018. https://doi.org/https://doi.org/10.1002/ppj2.20018

Yuan, J., Zhang, Y., Zheng, Z., Yao, W., Wang, W., & Guo, L. (2024). Grain crop yield prediction using machine learning based on UAV remote sensing: A systematic literature review. *Drones*, *8*(10). https://doi.org/10.3390/drones8100559

Zarr Development Team. (2024). *Zarr* (Version 2.18.2). Python Package Index. https://pypi.org/project/zarr/

DRAFT