Variational Autoencoders Capture Plant Biomolecular Stress Information

Anoop C. Patil[1*], Benny Jian Rong Sng[1,2*], Yu-Wei Chang[3*], Joana B. Pereira[4], Chua Nam-Hai[1,2,3], Rajani Sarojam[1,2], Gajendra Pratap Singh[1+], In-Cheol Jang[1,2,5+], and Giovanni Volpe[1,3+]

[1]Disruptive & Sustainable Technologies for Agricultural Precision, Singapore-MIT Alliance for Research and Technology, Singapore.
[2]Temasek Life Sciences Laboratory, National University of Singapore, Singapore.
[3]Department of Physics, University of Gothenburg, Gothenburg, Sweden.
[4]Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden.
[5]Department of Biological Sciences, National University of Singapore, Singapore.
*Contributed equally to this work. +Supervised this work.

1. Introduction

Plants are exposed to various stresses, including extreme light and shade conditions, and infections, which trigger complex physiological and biochemical changes [1]. These changes are mediated through dynamic adjustments in various biomolecules within the plant to adapt to adverse environments.

Conventional methods for studying these stress responses involve phenotyping or biochemical assays, which are limited in sensitivity and often miss subtle molecular shifts. Raman spectroscopy offers a non-destructive window into these biochemical changes, capturing spectral fingerprints from key biomolecules. However, the high-dimensional non-linear nature of Raman data poses challenges in interpreting these spectra.

We propose a deep learning-based unsupervised approach using Variational Autoencoder (VAE) to address this challenge. VAEs embed high-dimensional Raman spectra into a low-dimensional latent space, where stress-induced spectral changes naturally cluster, allowing intuitive visualization and interpretation of stress adaptation. In this paper, we focus on the ability of the latent space of the VAE to understand and categorize various stress conditions. The interpretability of the latent space and the automatic identification of stress-specific signatures are part of a larger work. We apply this framework to capture stress responses in plant species such as *Arabidopsis*, *Choy Sum*, and *Kai Lan* under light, shade avoidance, and bacterial stresses.

2. VAE for Plant Stress Monitoring

A Variational Autoencoder (VAE) is a generative model that compresses the input data into a latent space while learning to reconstruct the original input. The VAE consists of the following-

1. Encoder: Maps high-dimensional Raman spectra into a latent distribution $\mathbf{z}$ parameterized by $\boldsymbol{\mu}$ (mean) and $\boldsymbol{\sigma}$(variance).
2. Decoder: Reconstructs Raman spectra from latent space $\mathbf{z}$.
3. Objective: Minimize a loss function that balances reconstruction error and the Kullback-Leibler (KL) divergence, ensuring a continuous latent space. This combination encourages interpretable latent spaces where spectral shifts due to stress can be easily visualized.

2.1 Related work

Current techniques for analyzing plant stress responses often rely on manual processing of Raman spectra, which can be time-consuming and prone to human error. In contrast, the VAE framework automates this process, providing a more accurate and unbiased identification of stress-related molecular patterns. Our VAE approach addresses these gaps by

1. Learning latent spaces directly from raw Raman spectra as shown in Fig. 1.
2. Allowing unsupervised separation of stress responses at species, mutants, and time point levels.

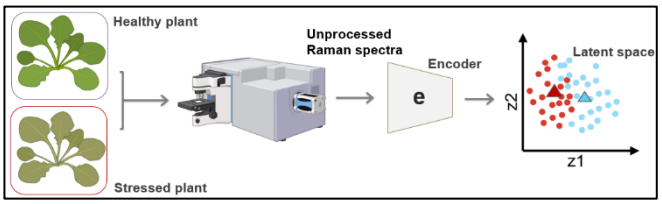

Fig. 1: VAE's encoder captures plant stress information from raw Raman spectra in the latent space.

2.2 Unsupervised Encoding of Light Stress Responses

We applied the VAE framework to decode light stress responses in Arabidopsis, Choy Sum, and Kai Lan plants exposed to white light, high light, low light, and shade conditions. Raman spectra were acquired from leaves under each condition. The VAE latent spaces illustrated in Fig. 2a showed clear separation between the four light conditions in wildtype plants. The median position shifted progressively from high light to white light, low light, and shade, indicating systematic biochemical changes. High light induced distinct molecular signatures characterized by increased carotenoid signals, while low light and shade conditions showed reduced carotenoid peaks, reflecting the plant's strategy to balance light capture and photoprotection. The clear separation of median positions for each plant species highlights the ability of our approach to capture molecular adaptations under different light stress conditions across multiple plant species.

2.3 Unsupervised Encoding of Shade Avoidance Stress Responses

We applied the VAE framework to decode shade avoidance stress [2] in wildtype Arabidopsis and two phytochrome mutants, *phyA-211* (PHYA gene knocked-out) and *phyB-9*^{BC} (PHYB gene knocked-out). Plants were grown under white light, moderate shade, and deep shade conditions, and Raman spectra were acquired from leaves. The VAE latent space showed clear separation between white light, moderate shade, and deep shade conditions in wildtype plants (Fig. 2b, left). The median position shifted upward with increasing shade severity, indicating progressive biochemical changes. In *phyA-211*, the separation between light conditions was more pronounced (Fig. 2b, center), reflecting PHYA's role in suppressing shade avoidance under low red to far-red (R:FR) light. In contrast, *phyB-9*^{BC} showed constitutive shade avoidance phenotypes under all light conditions, with white light and deep

shade spectra clustering together (Fig. 2b, right), corroborating PHYB's role in suppressing shade avoidance under high R:FR light.

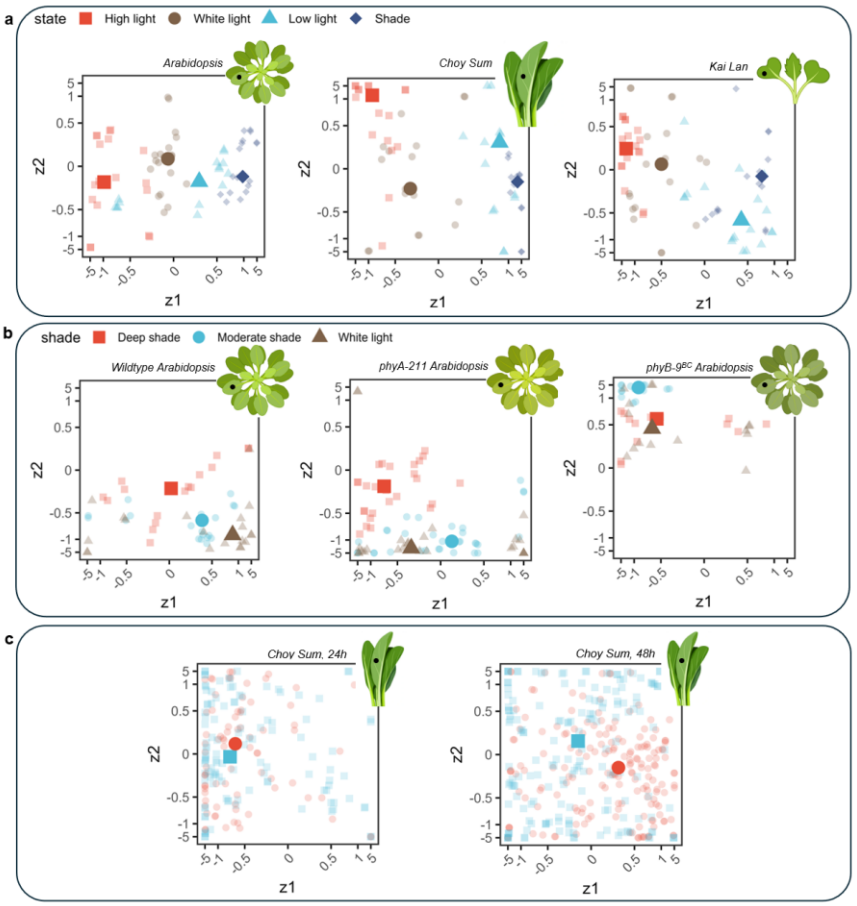

Fig. 2: VAE latent spaces for (a) light stress study, (b) shade avoidance stress study, and (c) bacterial stress study, across various plant species, mutants, and time points.

2.4 Unsupervised Encoding of Bacterial Stress Responses

For the bacterial stress study [3], wildtype Choy Sum plants were infiltrated with *Xanthomonas campestris pv. campestris* (*Xcc*) pathogens. Buffer-infiltrated plants served as controls at each time point. Raman spectra were collected at 24- and 48-hours post-infiltration. The VAE latent space for Choy Sum data at 24 hours showed slight separation between buffer- and pathogen-infiltrated plants (Fig. 2c, left). By the 48-hour mark (Fig. 2c, right), the separation increased, indicating a stronger infection response.

3. Conclusion

This study shows the capability of unsupervised VAEs for decoding plant stress responses from Raman spectra. By embedding spectra into interpretable latent spaces, VAEs can categorize gene-specific and species-specific stress signatures without requiring prior labels. This unsupervised approach reduces the need for human bias and prior knowledge, providing a scalable tool for real-time plant stress monitoring and early stress detection in crops. The ability to capture complex plant responses across species and stress types positions this workflow as a valuable tool for plant phenotyping and agricultural research.

Acknowledgments

The study was supported by (i) Disruptive \& Sustainable Technologies for Agricultural Precision (DiSTAP), Singapore-MIT Alliance for Research and Technology (SMART), (ii) Temasek Life Sciences Laboratory (TLL), National University of Singapore (NUS) and (iii) the National Research Foundation (NRF). Data curation for deep learning was carried out using *BRAPH2DiSTAP* software, a customized fork of the BRAPH 2.0 software (https://github.com/braph-software/BRAPH-2). The authors thank the researchers from the Genome \& Ecological Biology, Molecular Pathogenesis, and Developmental Biology groups at TLL for providing the Raman spectroscopy data for deep learning tasks reported in this work and for their valuable support throughout the study.

References

[1] Lew, T. T. S. et al. Species-independent analytical tools for next-generation agriculture. *Nature Plants* 6, 1408–1417 (2020).

[2] Sng, B. J. R. et al. Rapid metabolite response in leaf blade and petiole as a marker for shade avoidance syndrome. *Plant Methods* 16, 1–17 (2020).

[3] Chung, P. J., Seo, J. S., Ram, R. J., Sarojam, R. & Chua, N.-H. Rapid detection and quantification of plant innate immunity response using Raman spectroscopy. *Frontiers in Plant Science* 12, 746586 (2021).

Keywords:

Raman spectroscopy, Variational Autoencoders, Unsupervised learning, Plant stress, Deep learning, Precision agriculture.