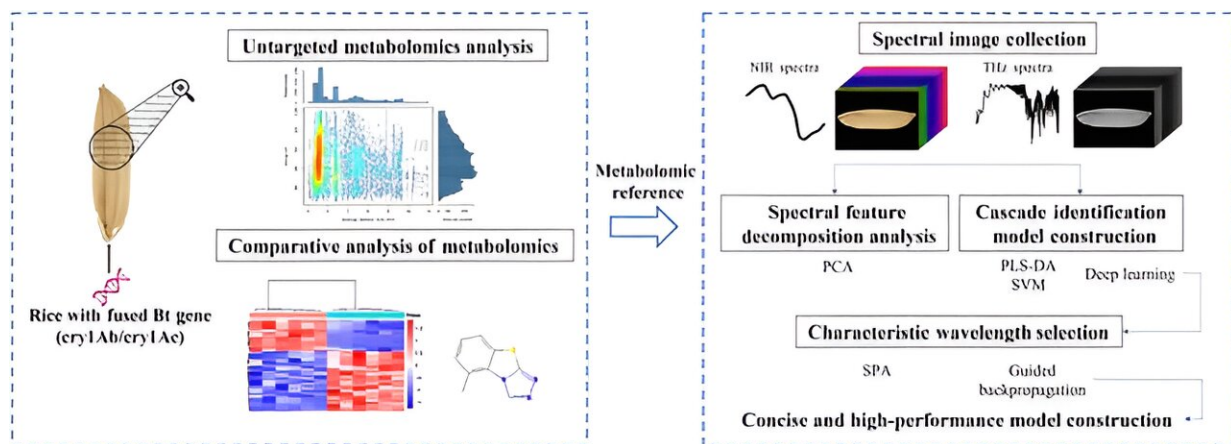# Novel spectral phenotyping and deep learning approach unveiled for GM rice seed detection

December 20 2023

Flowchart of untargeted metabolomics analysis, spectral data acquisition, and data analysis for discriminating GM rice seeds. Credit: *Plant Phenomics*

Cereals, crucial for food and biofuel, are increasingly using genetically modified (GM) technology in production to enhance resistance and nutrition. Despite biosafety concerns, precise detection methods like PCR are complex and expert-dependent. Advanced spectroscopic techniques, including near infrared and terahertz, are emerging as simpler alternatives for identifying GM organisms by analyzing spectral signatures.

Machine learning, especially deep learning algorithms, has significantly improved the accuracy of these methods, though challenges remain in

managing [high-dimensional data](#) and understanding the "black box" nature of deep learning. It is of great importance to refine these techniques for broader application and clearer interpretability.

In July 2023, *Plant Phenomics* published a research article titled "[Concise Cascade Methods for Transgenic Rice Seed Discrimination using Spectral Phenotyping](#)."

In this study, researchers first investigated the metabolic variability in rice seeds expressing the cry1Ab/cry1Ac gene. Metabolome analysis revealed significant differences between GM and non-GM rice varieties, with organic acids, lipids, and organ heterocyclic compounds being predominant. Spectral analysis showed that non-GM seeds generally had higher reflectance values than GM seeds.

The PCA of NIR spectra indicated a challenge in classification based solely on original data, highlighting the need for feature extraction. Terahertz spectra, although less visually separable, showed notable absorption peaks and variations among genotypes.

The CascadeSeed-1 model, using NIR and terahertz spectra, demonstrated superior accuracy in variety discrimination compared to other [machine learning](#) models. For GM status identification, the CascadeSeed-2 model showed high accuracy across different rice varieties, with terahertz spectra-based models generally outperforming NIR-based ones.

Wavelength selection methods were employed to refine the model, reducing redundant features. Although this resulted in some decrease in accuracy, it significantly enhanced the speed and efficiency of the model.

The study confirmed a correlation between metabolomic and spectral

analyses, with spectral features reflecting metabolic content. The cascade modeling approach efficiently recognized transgenic seeds from different genetic backgrounds, and therefore it outperforms traditional machine learning methods.

The guided back-propagation algorithm was effective in identifying characteristic wavelengths, correlating with specific metabolic changes.

In conclusion, this study developed a non-invasive, rapid method for identifying GM rice seeds using NIR and terahertz spectroscopy coupled with deep learning algorithms.

While slight accuracy trade-offs were observed with simplified models, this approach shows promise for enhanced real-world applicability in detecting GM organisms and conducting risk assessments.