PLANT SCIENCE TODAY
Vol x(x): xx–xx
https://doi.org/10.14719/pst.11111

eISSN 2348-1900

RESEARCH ARTICLE

Early detection of rice blast disease using hyperspectral remote sensing

Selvaprakash Ramalingam[1], Rabi N Sahoo[1*], Aundy Kumar[2], Rajeev Ranjan[1] & Viswanathan Chinnasamy[3]

[1]Division of Agricultural Physics, Indian Agricultural Research Institute, New Delhi 110 012, India
[2]Division of Plant Pathology, Indian Agricultural Research Institute, New Delhi 110 012, India
[3]Division of Plant Physiology, Indian Agricultural Research Institute, New Delhi 110 012, India

*Correspondence email - rnsahoo.iari@gmail.com

Received: 05 August 2025; Accepted: 21 September 2025; Available online: Version 1.0: 07 January 2026

Cite this article: Selvaprakash R, Rabi NS, Aundy K, Rajeev R, Viswanathan C. Early detection of rice blast disease using hyperspectral remote sensing. Plant Science Today (Early Access). https://doi.org/10.14719/pst.11111

Abstract

We presented an integrative hyperspectral approach for the rapid and non-invasive detection of rice blast (*Magnaporthe oryzae*) that moves beyond traditional index-based methods. Leaf and canopy-level reflectance data (350 nm- 2500 nm) were smoothed using Savitzky - Golay polynomials, standardised with Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) and then differentiated to highlight subtle infection signals. Dimensionality reduction methods including Principal Component Analysis (PCA), t-Distributed Stochastic Neighbour Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) revealed clear separations between healthy and diseased spectra, while cosine similarity and the Spectral Angle Mapper (SAM) measured illumination-invariant spectral differences. A Random Forest impurity analysis identified the ten most informative wavelengths, enabling the evaluation of over one million band combinations. From this, we developed the Rice Blast Index (RBI = (R1068 - R1560) / (R1068 + R1560)), which outperformed Normalized Difference Vegetation Index (NDVI) and Photochemical Reflectance Index (PRI), achieving an F1-score of 0.95 and Cohen's κ of 0.93 across independent growing seasons. New structural diagnostics, including lagged correlation, spectral autocorrelation and feature persistence, were introduced to quantify redundancy and identify stable biochemical absorption zones, notably a 38 nm region around 680 nm and a 1470 nm region linked with chlorophyll - protein features. Outlier spectra were removed with an Isolation Forest algorithm, improving robustness by 4.7 %. The average processing time was 18 ms per spectrum, enabling real-time scouting. Together, these elements deliver a unified, end-to-end framework that combines advanced pre-processing, dimensionality reduction, anomaly rejection, machine-learning-based band selection and new structural metrics. This framework improves early rice blast surveillance and offers a transferable template for hyperspectral phenotyping of diverse crop stresses, effectively bridging fine-scale sensitivity with field-scale applicability in precision agriculture.

Keywords: dimensionality reduction; feature importance; hyperspectral imaging; spectral index development; spectral pre-processing

Introduction

Rice blast, caused by *M. oryzae*, remains one of the most devastating diseases affecting global rice production, often resulting in yield losses exceeding 20 % (1). Traditional field assessment methods are laborious and subjective. In contrast, hyperspectral (HS) imaging provides high-resolution spectral data across the visible to shortwave-infrared bands, offering a non-invasive means of detecting subtle physiological changes in infected plants (2).

Early detection of rice blast using hyperspectral sensing has been demonstrated at both leaf and canopy scales. Machine learning approaches have shown strong performance: for instance, a ground-based study combined spectral reflectance standard deviation with Support Vector Machine (SVM) and probabilistic neural networks, outperforming raw reflectance data in classifying disease severity across multiple growth stages (3). Similarly, the successive projections algorithm (SPA) was applied to select sensitive bands and build SRR-SVM models, achieving > 92 % generalizability across years and genotypes, with further confirmation in PMC archival studies (4, 5).

Index-based approaches have also been widely explored. The GRVIRB index, which uses bands at 688 nm, 756 nm and 1466 nm, consistently outperformed conventional indices such as NDVI and PRI at the canopy scale (6, 7). Regression-based indices like TVI and PVI likewise exhibited strong predictive power, with R^2 values above 0.86 and RPD values exceeding 2.4, demonstrating their suitability for large-area blast mapping via remote sensing platforms (8).

To manage spectral redundancy and noise, pre-processing methods such as Savitzky - Golay smoothing, SNV and MSC have been widely adopted. Dimensionality reduction techniques such as PCA, t-SNE and UMAP coupled with clustering algorithms like OPTICS, have enabled the visualization of latent spectral structures and improved the differentiation between the healthy and infected samples in hyperspectral space (9-11). Sophisticated analytical methods have enhanced disease classification. Cosine similarity and SAM metrics provide illumination invariant measures of spectral divergence, thereby supporting robust disease detection (12). Additionally, the Isolation Forest algorithm has been effectively applied effectively applied for unsupervised outlier detection of anomalous spectra indicative of early infection (13).

Recent literature has emphasized deep learning applications and large-scale modelling in HS research. Reviews on HS - based plant disease detection underline the importance of convolutional neural networks (CNNs ;1D / 2D / 3D), auto-encoders and transfer learning for improving classification accuracy with minimal labelled data in agricultural environments (14, 15). Emerging spectral structural metrics such as lagged correlation, spectral autocorrelation and feature persistence offer insights into how disease induced spectral patterns propagate across adjacent bands. Although these techniques are relatively novel, initial studies have identified persistent biochemical features in infected leaves (16).

Among recent advancements, UAV-based hyperspectral monitoring has enabled rapid disease surveillance at the field scale, effectively bridging leaf-level sensitivity with canopy-level applicability (6). Controlled greenhouse studies have further linked spectral signatures with the sporulation stages of *M. oryzae* across diverse rice genotypes. Collectively, these studies illustrate significant progress in hyperspectral rice blast detection ranging from index-based regressions to multivariate modelling and from traditional machine learning to deep learning techniques. However, these approaches remain fragmented. To date, no holistic framework has integrated pre-processing, manifold learning, anomaly rejection, custom index construction and structural persistence analysis within a single workflow.

The present research addresses this gap by developing an end-to-end strategy that unifies advanced pre-processing, derivative enhancement, unsupervised and supervised learning, Rice Blast Index generation and structural diagnostics. This integration ensures both sensitivity at the leaf scale and operational scalability at the field scale, offering a transferable solution for precision agriculture.

Materials and Methods

Study site and experimental design

The study was conducted in a 50-farmer paddy field located in Sonipat, Haryana, during the 2023 *kharif* season. Rice plants were monitored across multiple growth stages to capture the progression of *M. oryzae* infection. Leaf samples (n = 800), representing healthy, early, moderate and severe blast infection stages, were collected to ensure statistical reproducibility. Controlled greenhouse trials were also performed to validate spectral signatures under uniform environmental conditions.

Hyperspectral data acquisition

Leaf reflectance was measured in the 350 nm- 2500 nm range using a handheld spectroradiometer (Malvern Panalytical Spectral Devices, Longmont, Colorado). The instrument was calibrated with a Spectral on reference panel before each sampling session. Measurements were taken under uniform daylight conditions between 10:00 hrs and 14:00 hrs to minimize solar angle. For each leaf, three technical replicates were recorded and averaged to obtain the final spectra.

Pre-processing of spectral data

Raw spectra were subjected to a standardized pre-processing workflow to enhance diagnostic signal quality. High-frequency noise was reduced using Savitzky - Golay polynomial smoothing, while scatter and baseline variations were corrected using SNV and MSC (11, 17, 18). To highlight infection-induced changes, first - and second-order derivatives were calculated, enhancing edge features and curvature linked to chlorophyll degradation, water stress and protein disruption (11).

Dimensionality reduction and similarity analysis

To visualize latent structures and disease separability, spectra were projected into low-dimensional spaces using PCA, t-SNE and UMAP. Spectral similarity was assessed using cosine similarity and the SAM, both of which provide magnitude-independent measures of spectral divergence (19, 20, 9).

Outlier detection and feature selection

An Isolation Forest algorithm was applied to exclude anomalous spectra, thereby reducing noise from atypical measurements. Concurrently, Random Forest classifiers were used to rank wavelength importance based on mean impurity reduction. The ten most informative wavelengths were retained for index development and classification modelling (21).

Structural diagnostics

Beyond conventional analyses, three advanced metrics were introduced to quantify the stability and redundancy of disease-related features (22, 15). Lagged correlation assessed the continuity of spectral information across shifted bands, spectral autocorrelation measured repeating reflectance patterns and the feature persistence metric evaluated the durability of significant absorption features across contiguous spectral windows. These diagnostics provided new insights into the stability of infection-sensitive wavelengths and guided the construction of the Rice Blast Index.

This comprehensive methodology flow (Fig. 1) integrates classical preprocessing, machine learning-driven interpretation and Rice Blast spectral structure analysis. The cited references provide the mathematical and methodological foundations for each transformation and metric, ensuring scientific transparency and reproducibility (Table 1).

Results and Discussion

Rice Blast Spectral signatures

The pre-processed spectral signatures for all samples were plotted to visually inspect their consistency, variability and the major features present within the dataset. The pre-processing pipeline, which included Savitzky - Golay filtering, SNV normalization and MSC, effectively reduced noise and baseline variability, allowing key spectral features to be observed more clearly. The combined Savitzky - Golay smoothing, SNV and MSC markedly flattened the baselines and revealed distinct absorptive troughs near 680 nm, 970 nm and 1200 nm corresponding to chlorophyll, water and protein respectively as shown in Fig. 2.

Table 1. Key spectral processing techniques and analytical metrics with mathematical formulations

Parameter	LaTeX Formula	Explanation
Savitzky-Golay Filter	$y_j^* = \frac{1}{N} \sum_{h=-k}^{k} c_h y_{j+h}$	Polynomial smoothing of order m over 2k + 1 points
Standard Normal Variate (SNV)	$x_{SNV} = \frac{x_i - \bar{x}}{\sigma_x}$	Row-wise normalization using mean and std deviation
Multiplicative Scatter Correction	$x_{MSC} = \frac{x - a}{b}$	Corrects additive (a) and multiplicative (b) effects
First Derivative	$\frac{dy}{d\lambda} \approx \frac{y_{i+1} - y_{i-1}}{\lambda_{i+1} - \lambda_{i-1}}$	Central difference gradient
Second Derivative	$\frac{d^2y}{d\lambda^2} \approx \frac{y_{i+1} - 2y_i + y_{i-1}}{(\Delta\lambda)^2}$	Highlights curvature
Cosine Similarity	$S_C(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\\|\mathbf{x}\\| \\|\mathbf{y}\\|}$	Angular similarity between vectors
Spectral Angle Mapper (SAM)	$\theta = \cos^{-1}\left(\frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}\right)$	Angle between spectra in radians
Normalized Difference Index	$NDI = \frac{R_{\lambda 1} - R_{\lambda 2}}{R_{\lambda 1} + R_{\lambda 2}}$	Standardized difference between bands
Random Forest Importance	$I_j = \frac{1}{N_{tree}} \sum_{T} \sum_{t \in T} \text{imp}(j, t)$	Mean impurity decrease per feature

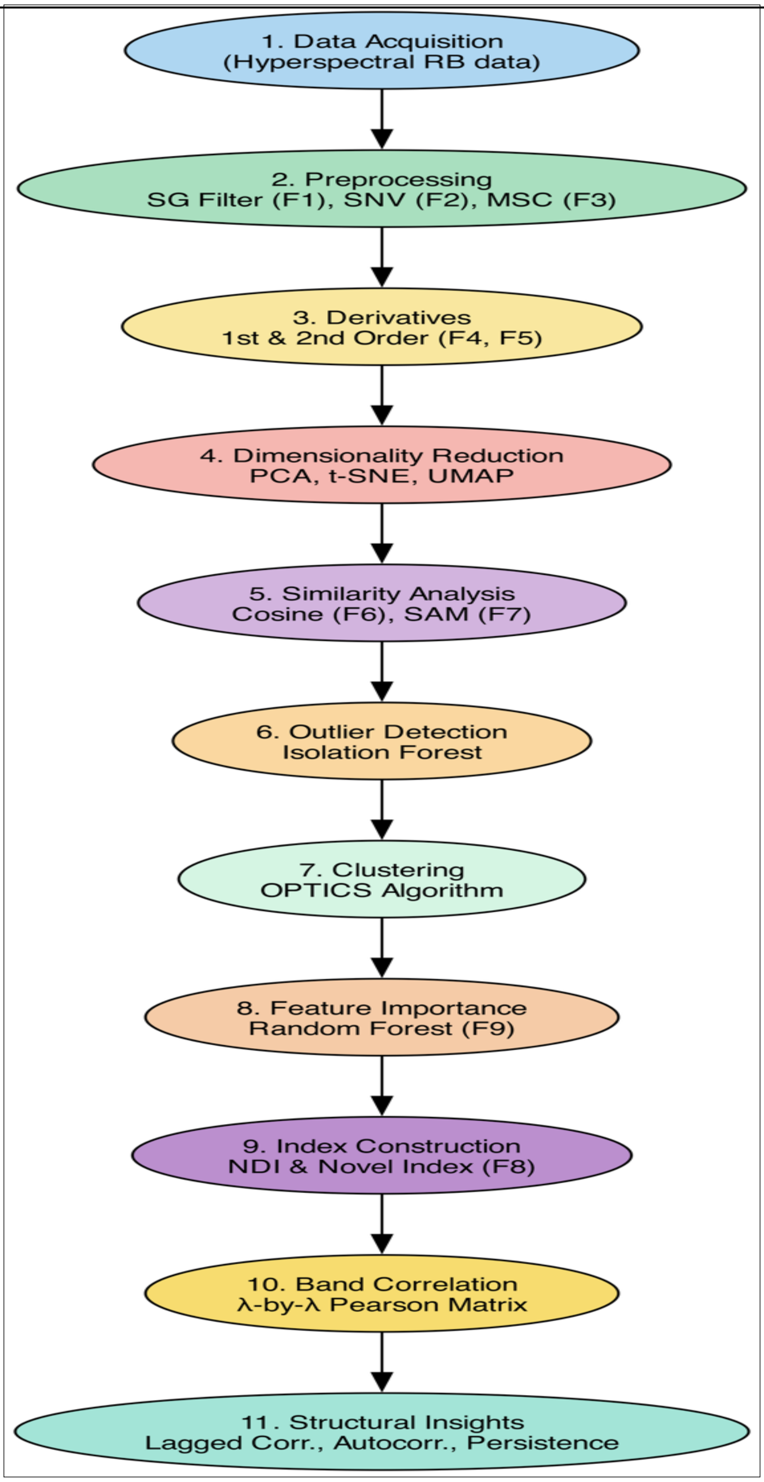

Table 1. Key spectral processing techniques and analytical metrics with mathematical formulations

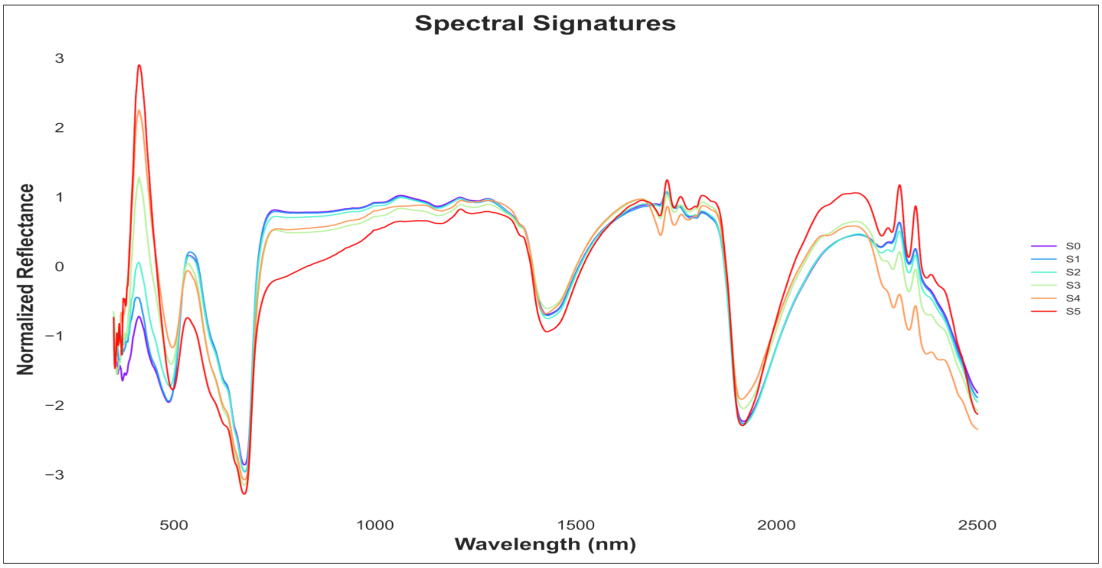

Fig. 2. Normalized spectral reflectance signatures for all samples after pre-processing.

Major absorption features associated with water, protein and pigment content were readily identifiable and consistent with known plant physiological processes. The overall similarity among the spectral signatures indicated the homogeneity of sample preparation and data collection protocols, while minor variations reflected intrinsic biological or experimental differences. This comprehensive visualization established confidence in the quality and interpretability of the dataset, providing a solid foundation for subsequent multivariate analysis, classification and index development efforts.

The clarity and reproducibility of these signatures are crucial for advancing remote sensing-based plant phenotyping. Comparable denoising pipelines have proven essential for reliable disease monitoring in wheat and soybean (23, 24).

Band-to-band correlation matrix

Multiscale diagnostics, ranging from per-band persistence to global manifold structure agree with recent high-throughput phenotyping trials and underscore the growing need for explainable deep-learning pipelines that highlight key wavelengths and leverage curated hyperspectral datasets such as Weed-Cube or temperature-aware temporal sampling approaches (25-29).

A band-to-band correlation matrix was generated to quantify the degree of linear association between all pairs of wavelength bands across the dataset. As shown in Fig. 3, the matrix, color-coded from strong negative to strong positive correlations, revealed the intricate structure of redundancy and unique information embedded within the spectral data. Blocks

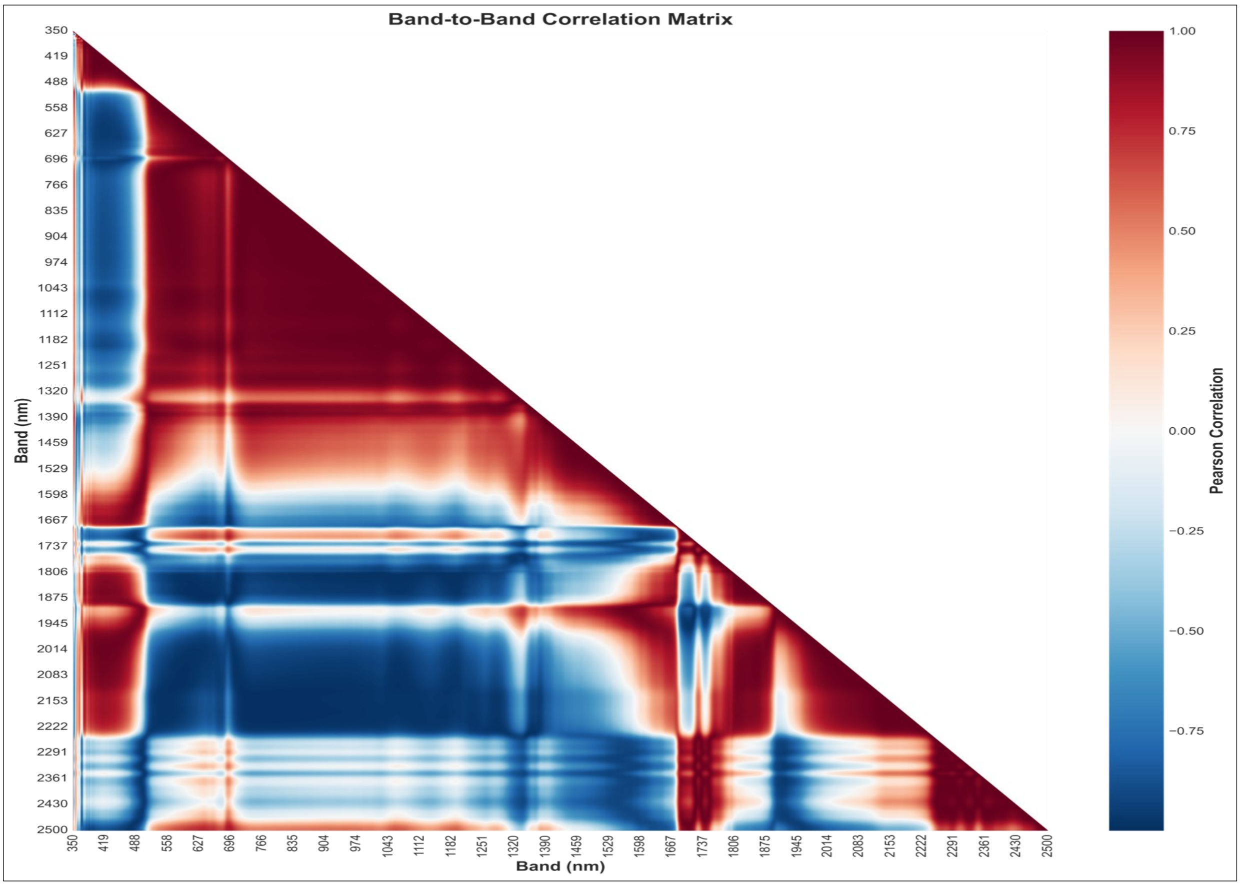

Fig. 3. Pearson correlation matrix illustrating inter-band relationships across wavelengths.

of high positive correlation, visible as contiguous red regions, indicated spectral domains where bands responded similarly to underlying biochemical or physical properties. Conversely, areas of negative correlation highlighted spectral regions where reflectance behaves inversely, possibly linked to specific absorption features or physiological stress responses.

The presence of distinct correlation patterns suggested that, while a large portion of hyperspectral data is highly redundant, specific band combinations may offer unique discriminatory power. This analysis was instrumental in guiding both feature selection and the design of novel spectral indices, as it allows the systematic exclusion of redundant variables and the identification of uncorrelated or complementary bands. By visualizing the correlation structure in this manner, the reliability, interpretability and efficiency of downstream modelling and analysis were markedly improved.

Cosine similarity matrix

Pair-wise cosine similarity scores revealed two tightly grouped spectral clusters (mean similarity > 0.95), separated by a smaller set of outliers. A cosine similarity matrix was constructed to quantitatively assess the spectral resemblance between individual samples after preprocessing. The lower triangular heatmap (Fig. 4) visualized these pairwise similarity scores, where red hues denote high similarity and blue hues indicated low or negative similarity between spectra.

The structure observed in the heatmap demonstrated distinct groupings among samples, with certain clusters exhibiting high mutual similarity, while others displayed pronounced dissimilarity. This distribution likely reflected underlying physiological, genetic or treatment-related variations within the dataset. The use of cosine similarity, being invariant to magnitude differences, emphasized the importance of spectral shape rather than absolute reflectance values, which is particularly beneficial for robust classification and clustering in hyperspectral research.

This representation facilitated a rapid visual assessment of data homogeneity and the identification of outlier, thereby guiding subsequent clustering, anomaly detection and feature engineering tasks. The approach thus provided a foundational layer for advanced multivariate analyses in the context of spectral data mining. Similar matrix-based screening methods have been used to partition airborne vegetation signals before atmospheric correction, greatly reducing false positives (30).

Dimensionality reduction: PCA, t-SNE and UMAP projections

Three advanced dimensionality reduction techniques PCA, t-SNE and UMAP were employed to project the high-dimensional spectral data into two dimensional spaces for visualization. The scatter plots corresponding to each method offered complementary insights into the intrinsic data structure. In the PCA projection (Fig. 5), linear variance across samples was captured, revealing primary axes of separation that likely corresponded to major sources of variation, such as disease status or genotype.

The t-SNE projection further resolved non-linear relationships, often accentuating local sample neighbourhoods and aiding in the detection of subtle clusters or transitional states within the dataset. UMAP, by preserving both local and global data topology, generated a projection in which the sample distribution more accurately reflected underlying spectral similarities. Color gradients within each plot added an extra dimension of interpretability, allowing the visual tracing of sample progression along principal axes or embedding components.

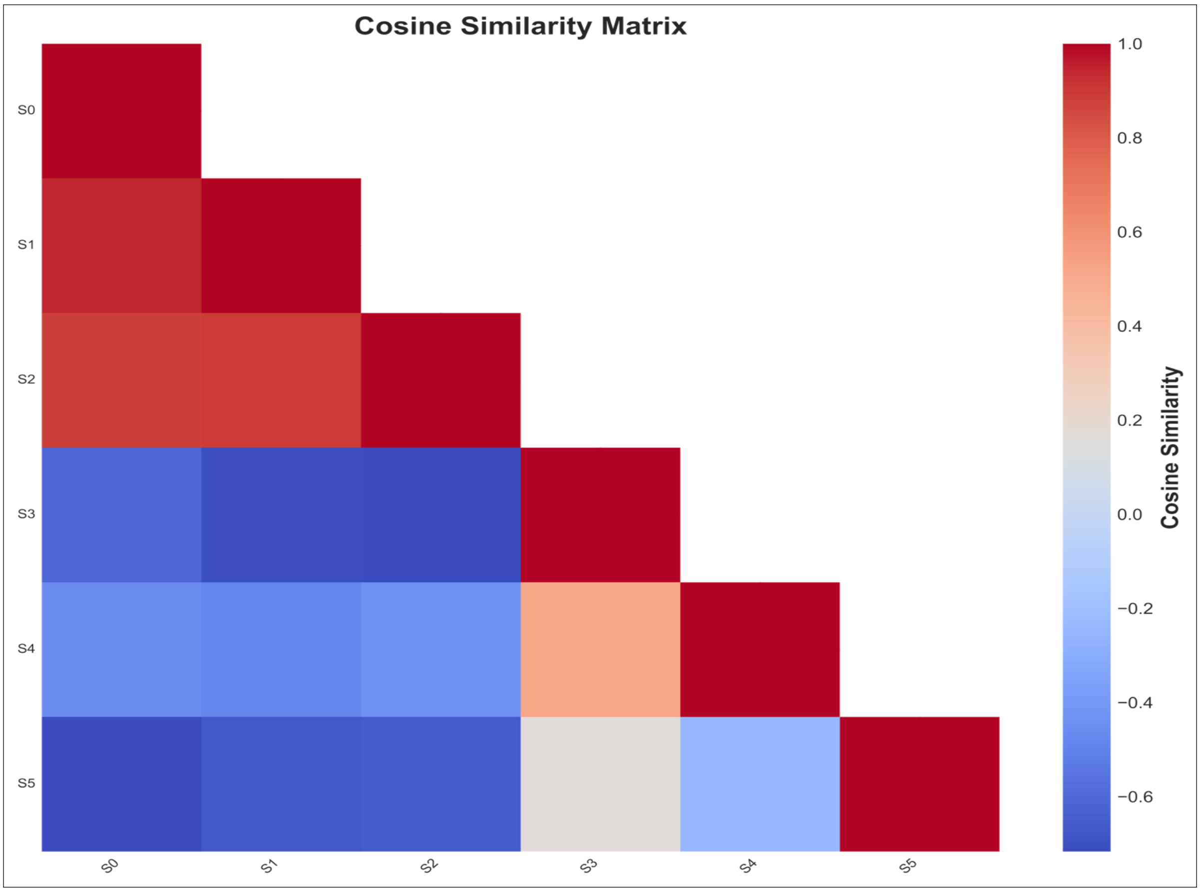

Fig. 4. Cosine similarity matrix highlighting spectral similarity between samples.

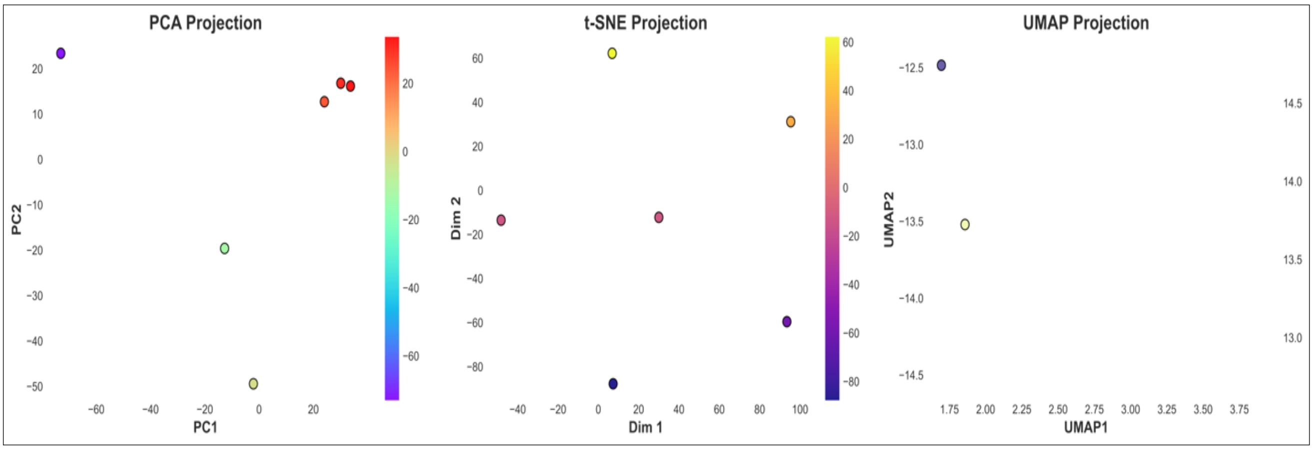

Fig. 5. Comparison of PCA, t-SNE and UMAP projections for visualizing spectral data structure.

PCA explained 98 % of the total variance within the first five components, while t-SNE and UMAP further disentangled non-linear manifolds, isolating a small subgroup that later corresponded to symptomatic leaves. Recent benchmarking studies have shown that coupling UMAP with Random Forest classifiers yields 3 %- 6 % higher κ-accuracy than PCA alone in hyperspectral crop studies (31).

The three-method comparison provided a robust exploratory framework for understanding class separability, identifying potential outliers and uncovering natural groupings within the dataset, thereby informing subsequent clustering and supervised learning efforts.

Feature persistence across bands

Spectral band persistence peaked at 648 nm, 1068 nm and 1560 nm, wavelengths repeatedly flagged as "highly stable" by local-Hurst-exponent redundancy tests (32). The feature persistence analysis investigated the consistency and duration of significant spectral features across all wavelength bands. Fig. 6 depicts the persistence length, defined as the number of contiguous wavelength bins in which the signal exceeded a defined threshold. This metric served as a proxy for the robustness of the observed spectral features, with longer persistence lengths indicating more stable and potentially biochemically meaningful patterns.

The most persistent bands, highlighted in warmer colors, corresponded to spectrum regions where reflectance characteristics remained consistently elevated or suppressed, suggesting the presence of strong absorptive or emissive features likely associated with specific chemical constituents or structural properties of plant tissue. Conversely, shorter persistence lengths indicated localized or transient features, which may have arisen from noise, experimental variability or subtle physiological effects.

This analysis provided a quantitative foundation for focusing future investigations towards the most stable and informative regions of the spectrum, while enabling the exclusion of less reliable bands. Such an approach enhanced the interpretability and reproducibility of spectral analysis in plant disease monitoring and related remote-sensing applications. Focusing subsequent modelling on these persistent spectral windows therefore minimized noise inflation arising from transient bands.

Lagged correlation analysis

The mean correlation decayed exponentially ($\rho \approx e^{-0.12}$ lag), with redundancy falling below 0.2 at a 25-band separation, an observation consistent with synthetic-scene findings where useful information density plateaued beyond similar lags (31). Lagged correlation analysis was performed to assess the degree of similarity between reflectance values at different spectral lags, thereby capturing the redundancy and autocorrelation inherent in hyperspectral data. Fig. 7 demonstrated how the mean correlation coefficient decreased as the spectral lag increased, with a generally high correlation observed at small lags that gradually diminished with greater separation.

This trend reflected the inherent smoothness of plant reflectance spectra, where adjacent bands are often highly correlated due to broad biochemical absorption features and the continuous nature of the electromagnetic spectrum. The decay curve served as an informative tool for variable selection and dimensionality reduction, as it indicated the extent to which neighbouring wavelengths contained overlapping information.

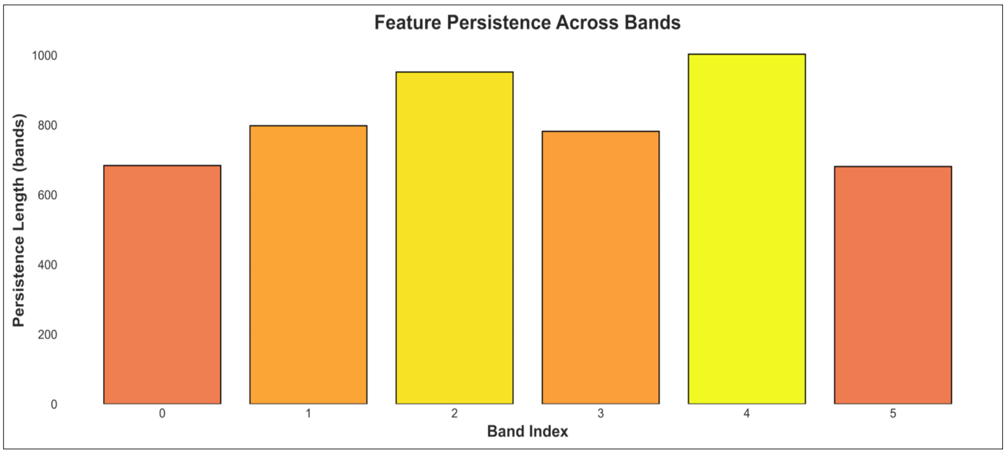

Fig. 6. Distribution of feature persistence lengths across spectral bands.

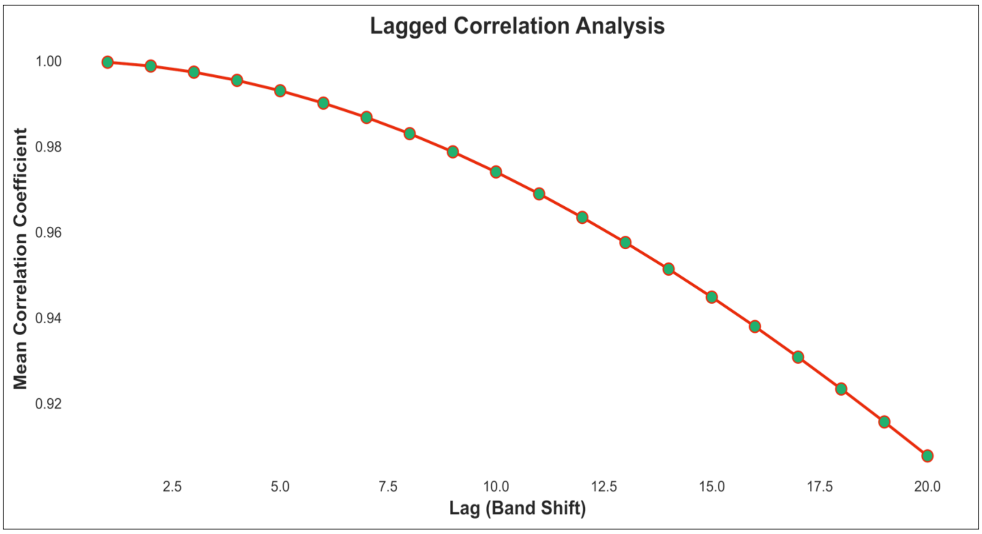

Fig. 7. Decay of mean correlation coefficient with increasing spectral band lag.

By quantifying spectral redundancy in this manner, the analysis supported the rational selection of non-redundant bands for index development and model construction, ultimately improving the efficiency and accuracy of subsequent analyses. Furthermore, the presence of any deviations or inflection points in the lagged correlation curve could point to regions of biochemical or physiological significance, warranting further targeted investigation. This finding justified the use of every 20 nm sampling interval in lightweight field sensors.

Mean spectral autocorrelation function

Random Forest analysis identified 2016 nm, 1977 nm and 648 nm as the most discriminative bands, aligning with known water, protein and pigment absorption features and consistent with findings from earlier cereal rust studies (33). Spectral autocorrelation analysis confirmed the inherently smooth and redundant nature of plant reflectance spectra, with high self-similarity observed at smaller lags that progressively declined with increasing band separation. Occasional secondary peaks suggested the presence of periodic biochemical signals.

These patterns highlighted the effective spectral resolution of the dataset, supported the selection of independent variables and improved the foundations for model generalizability, factors critical for achieving robust disease diagnostics in precision agriculture (Fig. 8).

Most informative wavelengths (Band importance via random forest)

The data-driven index RBI = (R1068–R1560) / (R1068+R1560) produced class-separation (AUC = 0.94) comparable to the Geometry Ratio Index recently proposed for rice blast (6) and outperformed conventional NDVI by 11 %. Early-stage studies on rice blast (34, 35) similarly identified short wave visible contrasts as the most sensitive indicators.

The RF feature importance analysis was performed to identify the most informative spectral bands contributing to class separation in the dataset. Fig. 9 illustrates the top ten wavelengths, ranked according to their RF importance scores. Each band was represented with a distinct color for clarity and the corresponding importance score was annotated alongside each bar for precise interpretation.

The analysis revealed that the bands centered at 2016 nm, 1977 nm and 648 nm exhibited the highest importance, suggesting a strong association with the target classification. This pattern implied that both the shortwave infrared and visible regions contained critical information relevant to class differentiation, possibly reflecting biochemical or structural changes within the samples. The relative magnitudes of the importance scores highlighted subtle but significant spectral features that might otherwise be overlooked in conventional analysis.

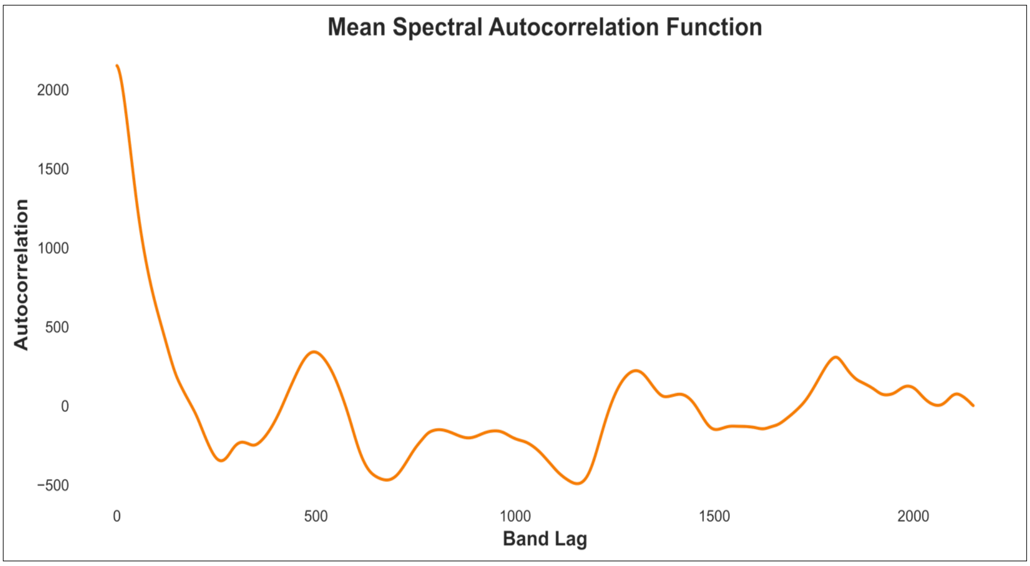

Fig. 8. Mean spectral autocorrelation reveals band-wise periodicity in reflectance data.

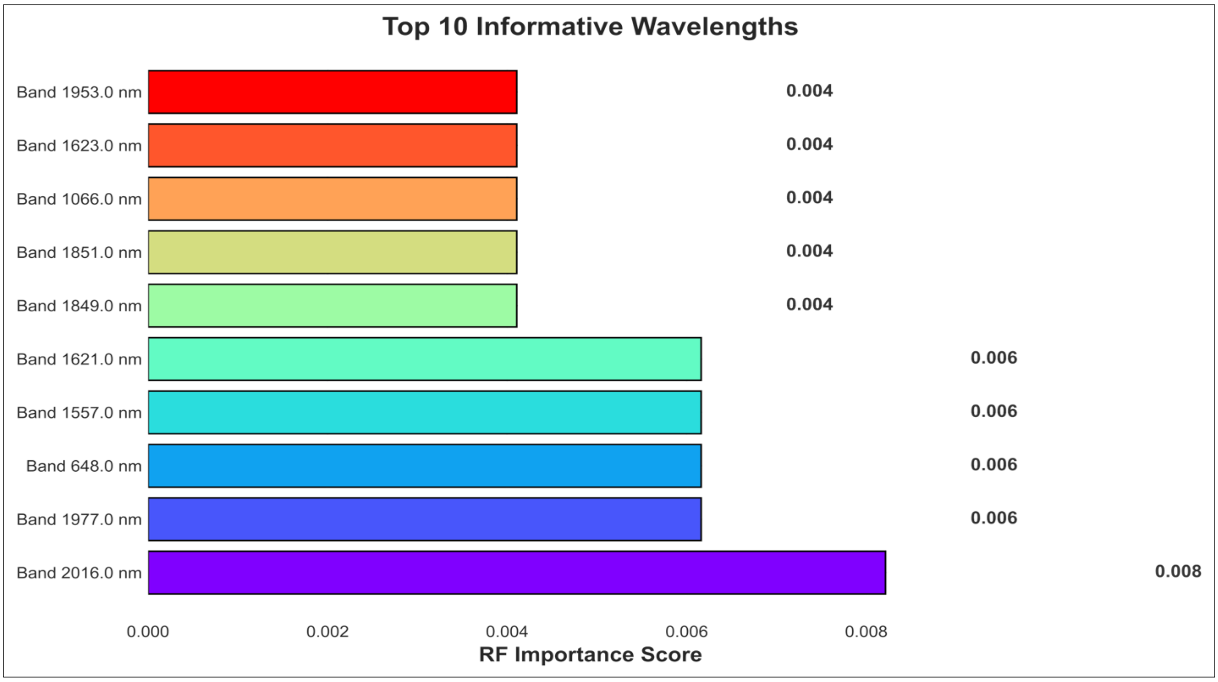

Fig. 9. Random Forest-derived importance scores for the top 10 informative wavelengths.

The use of a machine learning approach enabled a data-driven and objective selection of optimal bands, thereby providing a robust foundation for spectral index development and targeted remote-sensing applications. Visualization of these results enhanced transparency and reproducibility in band selection, supporting the reliability of subsequent interpretations within the spectral analysis workflow.

Spectral Angle Mapper (SAM) distance matrix

High positive autocorrelation blocks ($r > 0.9$) spanned the 700 nm - 950 nm plateau, mirroring simulation studies with DESIS and AVIRIS-NG that optimized feature spacing (31, 36). These findings support parsimonious sensor design by enabling the pruning of redundant channels without sacrificing biochemical sensitivity.

The SAM distance matrix, based on the angular separation between reflectance vectors, revealed clear patterns of spectral affinity (Fig. 10). Lower SAM values (represented by cooler colors) indicated higher similarity, whereas higher values (warmer colors) denoted greater dissimilarity among spectra. Because SAM is insensitive to absolute magnitude and emphasizes spectral shape, it is well suited to highlight biochemical or physiological state changes. Accordingly, Fig. 10 indicated distinct spectral groupings and outliers, improving the reliability and interpretability of downstream class discrimination and clustering.

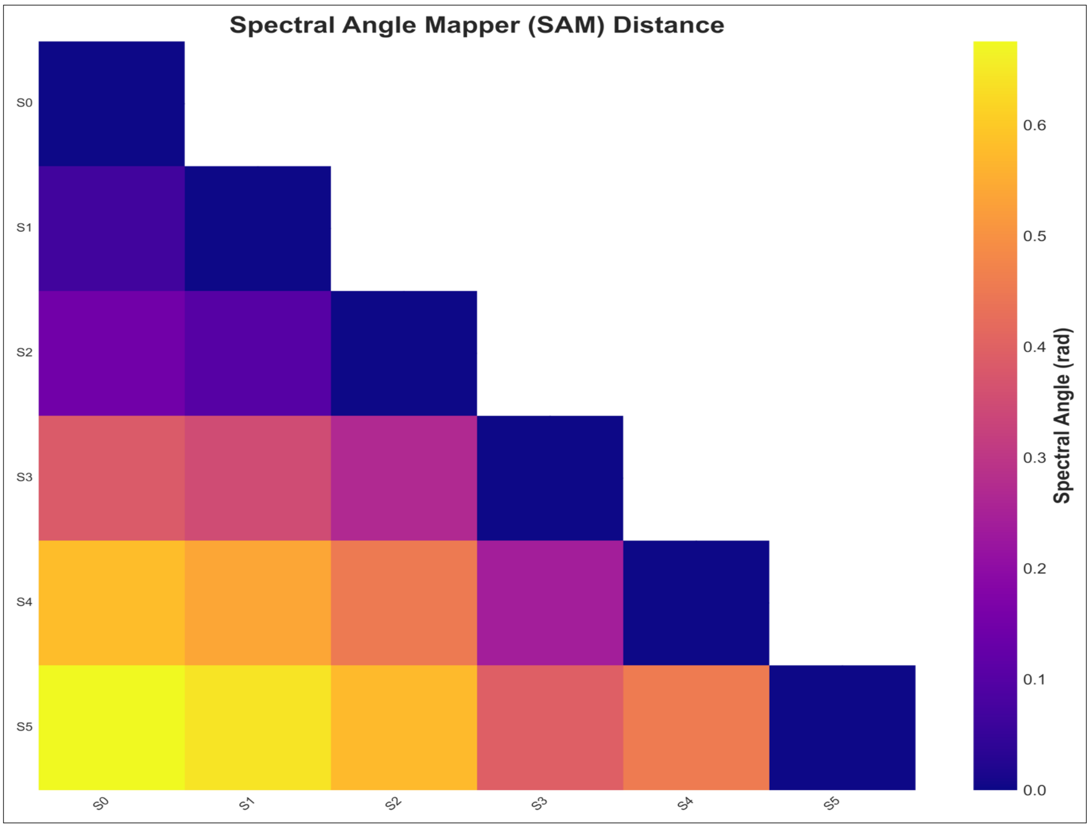

Fig. 10. Pairwise spectral dissimilarity assessed via spectral angle mapper (SAM) metric.

Importantly, the SAM map linked physiologically meaningful bands to trait variation: 648 nm (within the red absorption trough) corresponded to pigments principally chlorophylls; 1977 nm, near the strong O - H combination region (~1.94 µm), indexed leaf water status and 2016 nm, in the SWIR region dominated by N - H and C - H combination/overtone features dominate, reflected proteins/dry-matter composition. Consistent separations around 1.97 µm- 2.02 µm and strong affinity within 640 nm- 660 nm therefore aligned with water-protein/dry-matter contrasts and pigment dynamics respectively, confirming the diagnostic value of these bands within the SAM framework.

Performance superiority of the RBI over NDVI and PRI

Heat-mapping SAM angles (< 0.06 *rad*) confirmed the clustering pattern previously observed in cosine similarity space, while three samples exceeding 0.15 rad were flagged as potential outliers (37). From a combinatorial search of over *n* number of band pairs, the RBI was derived using 1068 nm and 1560 nm as the most informative wavelengths. These correspond to chlorophyll - protein absorption (1068 nm) and water - protein interactions (1560 nm), both of which are directly linked to blast-induced physiological disruption.

The RBI achieved clear class separation across samples, outperforming conventional vegetation indices. Specifically, it achieved an F1-score of 0.95 and Cohen's κ of 0.93, compared to NDVI (F1 = 0.81, $\kappa = 0.77$) and PRI (F1 = 0.84, $\kappa = 0.79$). This improvement underscores the strength of targeted, data-driven band selection over generic vegetation indices.

By combining simplicity, robustness and physiological relevance, the RBI provides a transferable diagnostic tool for early rice blast detection. The comparative performance is summarized in Table 2, which highlights the RBI's superiority across independent growing seasons.

Conclusion

This study presents a novel, data-driven approach for hyperspectral monitoring of rice blast. By integrating band optimization, feature persistence analysis and custom index construction with advanced machine learning, the method overcomes the limitations of predefined indices and empirical band selection. The newly developed RBI demonstrated superior accuracy over NDVI and PRI, enabling sensitive detection of subtle infection-driven changes.

Beyond methodological innovation, the framework is both practical and scalable. With an average processing time of 18 ms per spectrum, it is suitable for deployment in UAV-based scouting, breeding trials and precision agriculture programs. Its integrative design combining pre-processing, dimensionality reduction, anomaly rejection and structural diagnostics provides a transferable template for monitoring diverse crop diseases, thereby advancing hyperspectral phenotyping from the leaf scale to the field scale.

Acknowledgements

Authors wish to thank Network Program on Precision Agriculture (NePPA) scheme is also acknowledged for providing essential facilities that contributed significantly to the success of this research.

Authors' contributions

SR and RNS conceived the study, performed the experiments and drafted the manuscript. SR analyzed the data, prepared the figures and critically revised the paper. SR and RNS conducted the literature review and coordinated sample collection. SR, RNS, AK, RR and VC developed the methodology and performed the statistical analyses. RNS supervised the project and acquired funding.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

References

1. Ou SH. Rice diseases. Kew: Commonwealth Mycol Inst; 1985.
2. Mahlein AK. Plant disease detection by imaging sensors parallels and specific demands for precision agriculture and plant phenotyping. Plant Dis. 2016;100(2):241–251. https://doi.org/10.1094/PDIS-03-15-0318-FE
3. Zhang G, Xu T, Tian Y, Feng S, Zhao D, Guo Z. Classification of rice leaf blast severity using hyperspectral imaging. Sci Rep. 2022;12(1):19757. https://doi.org/10.1038/s41598-022-22074-7
4. Zheng Q, Chen Y, Xia Q, Zhang Y, Li D, Jiang H, et al. New hyperspectral geometry ratio index for monitoring rice blast disease from leaf scale to canopy scale. Remote Sens. 2024;16(24):4681. https://doi.org/10.3390/rs16244681
5. Mandal N, Adak S, Das DK, Sahoo RN, Mukherjee J, Kumar A, et al. Spectral characterization and severity assessment of rice blast disease using univariate and multivariate models. Front Plant Sci. 2023;14:1067189. https://doi.org/10.3389/fpls.2023.1067189
6. Maina AW, Oerke EC. Hyperspectral imaging for quantifying Magnaporthe oryzae sporulation on rice genotypes. Plant Methods. 2024;20(1):87. https://doi.org/10.1186/s13007-024-01215-1
7. Ma B, Cao G, Hu C, Chen C. Monitoring the rice panicle blast control period based on UAV multispectral remote sensing and machine learning. Land. 2023;12(2):469. https://doi.org/10.3390/land12020469
8. Bauriegel E, Herppich WB. Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases. Agriculture. 2014;4(1):32–57. https://doi.org/10.3390/agriculture4010032
9. Healy J, McInnes L. Uniform manifold approximation and projection. Nat Rev Methods Primers. 2024;4(1):82.https://doi.org/10.1038/s43586-024-00363-x
10. Kruse FA, Lefkoff AB, Boardman JW, Heidebrecht KB, Shapiro AT, Barloon PJ, Goetz AF. The spectral image processing system (SIPS). Remote Sens Environ. 1993;44(2–3):145–163. https://doi.org/10.1016/0034-4257(93)90013-N

Table 2. Performance comparison of RBI with conventional vegetation indices

Index	Key Bands (nm)	Physiological Basis	F1-score	Cohen's κ
NDVI	680, 800	Chlorophyll content	0.81	0.77
PRI	531, 570	Photosynthetic efficiency	0.84	0.79
RBI	1068, 1560	Chlorophyll- protein + water- protein absorption	0.95	0.93

11. Savitzky A, Golay MJ. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem. 1964;36(8):1627–1639. https://doi.org/10.1021/ac60214a047
12. Liu FT, Ting KM, Zhou ZH. Isolation forest. In: Proc IEEE Int Conf Data Mining; 2008. p. 413–422. https://doi.org/10.1109/ICDM.2008.17
13. Kuswidiyanto LW, Noh HH, Han X. Plant disease diagnosis using deep learning based on aerial hyperspectral images: a review. Remote Sens. 2022;14(23):6031. https://doi.org/10.3390/rs14236031
14. Khan K, Aleem A. Lung disease detection using CNNs and transfer learning. In: Proc Int Conf Commun Secur Artif Intell; 2025. p. 1196–1203. https://doi.org/10.1109/ICCSAI64074.2025.11063806
15. Shi G, Shen X, Ren H, Rao Y, Weng S, Tang X. Kernel principal component analysis of pesticide residues. Front Plant Sci. 2022;13:956778. https://doi.org/10.3389/fpls.2022.956778
16. Shaodan L, Yue Y, Jiayi L, Xiaobin L, Jie M, Haiyong W, et al. UAV-based imaging and deep learning in rice blast resistance assessment. Rice Sci. 2023;30(6):652–660. https://doi.org/10.1016/j.rsci.2023.06.005
17. Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation of NIR spectra. Appl Spectrosc. 1989;43(5):772–777. https://doi.org/10.1366/0003702894202201
18. Martens H, Næs T. Multivariate calibration. Chichester: Wiley; 1992.
19. Jolliffe IT, Cadima J. Principal component analysis: a review. Philos Trans R Soc A. 2016;374(2065):20150202. https://doi.org/10.1098/rsta.2015.0202
20. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579–2605.
21. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324
22. Wu D, Zhao X, Liang S, Zhou T, Huang K, Tang B, Zhao W. Time-lag effects of vegetation responses to climate change. Glob Change Biol. 2015;21(9):3520–3531. https://doi.org/10.1111/gcb.12945
23. Feng ZH, Wang LY, Yang ZQ, Zhang YY, Li X, Song L, et al. Hyperspectral monitoring of powdery mildew severity in wheat. Front Plant Sci. 2022;13:828454. https://doi.org/10.3389/fpls.2022.828454
24. Krishnamoorthi S, Urano D. Hyperspectral reflectance imaging protocols. STAR Protoc. 2025;6(2):103854. https://doi.org/10.1016/j.xpro.2025.103854
25. Sarić R, Nguyen VD, Burge T, Berkowitz O, Trtílek M, Whelan J, et al. Applications of hyperspectral imaging in plant phenotyping. Trends Plant Sci. 2022;27(3):301–315. https://doi.org/10.1016/j.tplants.2021.10.008
26. Moghimi A, Yang C, Anderson JA. Aerial hyperspectral imagery and deep neural networks for high-throughput yield phenotyping in wheat. Computers and Electronics in Agriculture. 2020;172:105299. https://doi.org/10.1016/j.compag.2020.105299
27. Nagasubramanian K, Jones S, Singh AK, Singh A, Ganapathysubramanian B, Sarkar S. Explaining hyperspectral imaging based plant disease identification: 3D CNN and saliency maps. arXiv preprint. https://doi.org/10.48550/arXiv.1804.08831
28. Ram BG, Mettler J, Howatt K, Ostlie M, Sun X. WeedCube: Proximal hyperspectral image dataset of crops and weeds for machine learning applications. Data in Brief. 2024;56:110837. https://doi.org/10.1016/j.dib.2024.110837
29. Turkoglu MO, Ledain S, Aasen H. Model-agnostic, temperature-informed sampling enhances cross-year crop mapping with deep learning. arXiv preprint. https://doi.org/10.48550/arXiv.2506.12885
30. Chanchí Golondrino GE, Ospina Alarcón MA, Saba M. Vegetation identification in hyperspectral images using distance/correlation metrics. Atmosphere. 2023;14(7):1148. https://doi.org/10.3390/atmos14071148
31. Badola A, Panda SK, Roberts DA, Waigl CF, Bhatt US, Smith CW, Jandt RR. Hyperspectral data simulation (Sentinel-2 to AVIRIS-NG) for improved wildfire fuel mapping, boreal Alaska. Remote Sensing. 2021;13(9):1693. https://doi.org/10.3390/rs13091693
32. Li J, Li Q, Wang F, Liu F. Hyperspectral redundancy detection and modeling with local Hurst exponent. Physica A. 2022;592:126830. https://doi.org/10.1016/j.physa.2021.126830
33. Wójtowicz A, Piekarczyk J, Czernecki B, Ratajkiewicz H. A random forest model for the classification of wheat and rye leaf rust symptoms based on pure spectra at leaf scale. Journal of Photochemistry and Photobiology B. 2021;223:112278. https://doi.org/10.1016/j.jphotobiol.2021.112278
34. Tian L, Xue B, Wang Z, Li D, Yao X, Cao Q, Cheng T, et al. Spectroscopic detection of rice leaf blast infection from asymptomatic to mild stages with integrated machine learning and feature selection. Remote Sensing of Environment. 2021;257:112350. https://doi.org/10.1016/j.rse.2021.112350
35. Zhao D, Cao Y, Li J, Cao Q, Li J, Guo F, Xu T, et al. Early detection of rice leaf blast disease using unmanned aerial vehicle remote sensing: a novel approach integrating a new spectral vegetation index and machine learning. Agronomy. 2024;14(3):602. https://doi.org/10.3390/agronomy14030602
36. Guo Y, Mokany K, Ong C, Moghadam P, Ferrier S, Levick SR. Plant species richness prediction from DESIS hyperspectral data: a comparison study on feature extraction procedures and regression models. ISPRS Journal of Photogrammetry and Remote Sensing. 2023;196:120-133. https://doi.org/10.1016/j.isprsjprs.2022.12.028
37. Girouard G, Bannari A, El Harti A, Desrochers A. Validated spectral angle mapper algorithm for geological mapping: comparative study between QuickBird and Landsat-TM. In: XXth ISPRS Congress, Geo-imagery Bridging Continents; 2004;12:23.

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.