Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 3
Open Access

Deep Spectral Signature Analysis and Feature Selection for Non-Destructive Curcumin Quantification in Turmeric Rhizomes from Hyperspectral Data

Mr. Sarfaraz Pathan [1], Dr. Sanjay Azade, [2] ,Deepali V. Sawane [3,] Mansur Shaikh [4,] Shabeena Khan [5]

Dr. G.Y. Pathrikar College of Computer Science & Information Technology, MGM University, Chhatrapati Sambhajinagar, Maharashtra,India. safaraz.ip@gmail.com

Cite this paper as: Mr. Sarfaraz Pathan, Dr. Sanjay Azade, Deepali V. Sawane, Mansur Shaikh, Shabeena Khan (2024) Deep Spectral Signature Analysis and Feature Selection for Non-Destructive Curcumin Quantification in Turmeric Rhizomes from Hyperspectral Data. *Frontiers in Health Informatics,* (3), 11486-11497

Abstract:

This study presents a spectral signature analysis and band selection framework for the non-destructive quantification of curcumin in fresh turmeric rhizomes using hyperspectral imaging (HSI). The analysis involves acquiring reflectance spectra across 200+ contiguous bands ranging from 388.9 nm to 1026.0 nm and identifying the most informative wavelengths associated with curcumin concentration. Mathematical feature selection techniques, including mutual information (MI), successive projections algorithm (SPA), and a hybrid MI-correlation-based ranking method, were applied to reduce spectral dimensionality.

Among these, the hybrid method demonstrated the best trade-off between band relevance and diversity. It consistently selected the chemically significant 425 nm band, as validated by HPLC, while enhancing deep learning model performance with lower RMSE. These findings make hybrid band selection a key enabler for efficient and interpretable 3D-CNN-based curcumin prediction.

Keywords: Hyperspectral Imaging, Band Selection, Curcumin, Turmeric Rhizomes, Spectral Signature, Dimensionality Reduction, Feature Engineering

Introduction:

Curcumin, the primary bioactive compound in turmeric (*Curcuma longa*), is valued for its pharmacological properties. Traditional methods for curcumin quantification such as HPLC are accurate but destructive and time-consuming. Hyperspectral imaging (HSI), which combines spatial and spectral information, offers a non-destructive alternative [15, 29]. This study explores the spectral signature of turmeric rhizomes and proposes an optimized band selection pipeline for curcumin prediction.

Materials and Methods:

Data Acquisition:

HSI data of fresh turmeric rhizomes were collected using a HSI camera covering wavelengths from 388.9 nm to 1026.0 nm. Each hyperspectral cube consisted of spatial dimensions (rows, columns) and spectral bands.

Reference Measurements:

Curcumin concentrations for selected regions of interest (ROIs) were measured using High-Performance Liquid Chromatography (HPLC) and used as ground truth. And P1, P2, P3, and Sample1 HIS images were extracted using ROI. Reflectance data were standardized using Z-score normalization.

Spectral Signature Extraction:

A spectral signature is the unique pattern of reflectance (or radiance) across different wavelengths for a specific material or object [66]. In the context of hyperspectral imaging (HSI), each pixel contains a full spectrum, and this spectral information can be used to identify or quantify materials like curcumin in turmeric [41].

Each turmeric rhizome pixel reflects light differently at each wavelength. These reflectance values, when plotted across the spectral range (e.g., 388.9 nm – 1026.0 nm), form the spectral signature. Variations in this signature indicate chemical or structural differences, such as varying curcumin concentrations.

Let the hyperspectral cube be denoted as:

$$H(x, y, \lambda) \in \mathbb{R}^{M \times N \times B}$$

Where:

x,,y are spatial coordinates (rows and columns),
λ denotes the wavelength index (1 to BB, e.g., 200+ bands),
H(x,y,λ) is the reflectance value at position (x,y) and band λ.

Region of Interest (ROI):

To extract meaningful data, we define a Region of Interest (ROI) — a manually selected area on the turmeric rhizome image known to contain curcumin.

Let RR be the set of all pixels in the ROI.

The average spectral signature $s(\lambda)$ over a ROI is:

$$s(\lambda) = \frac{1}{|R|} \sum_{(x,y) \in R} H\,(x, y, \lambda)$$

This averages the reflectance across all pixels in the ROI for each band λ, producing a 1D vector of length B — the mean spectral signature.

Preprocessing:

Raw spectra were smoothed using Savitzky-Golay filtering and normalized using standard score normalization (Z-score):

1. **Smoothing:**

Apply Savitzky-Golay filter to smooth spectral curves:

$$ssmooth(\lambda)=SG\_filter(s(\lambda))$$

2. **Normalization (Z-Score):**

$$s_{norm}(\lambda) = \frac{s(\lambda) - \mu}{\sigma}$$

where:

μ: mean of spectral signature

σ: standard deviation

This makes spectra comparable across samples, focusing analysis on shape rather than magnitude [33].

Band Selection Methods:

Mutual Information (MI):

Mutual Information between each band λ_i and the label y (curcumin concentration) was computed:

$$MI(\lambda_i, y) = \sum_{\lambda_i, y} p\,(\lambda_i, y) \log\left(\frac{p(\lambda_i, y)}{p(\lambda_i)p(y)}\right)$$

Optionally, Normalized Mutual Information (NMI) is computed as [63]**:**

$$NMI(\lambda_i, y) = \frac{2 \cdot MI(\lambda_i, y)}{H(\lambda_i) + H(y)}$$

Successive Projections Algorithm (SPA):

SPA selects bands iteratively:

$$\lambda_k = \underset{j}{\operatorname{argmax}} \left\| (I - P_{\Lambda_{k-1}}) s_j \right\|_2$$

where $P_{\Lambda_{k-1}}$ is the projection matrix from previously selected vectors.

Hybrid Ranking Score:

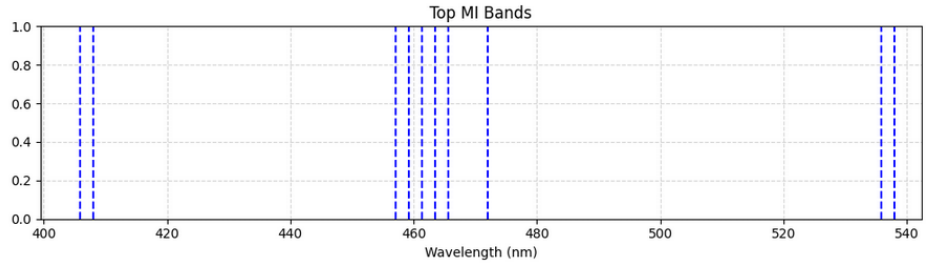
$$H(\lambda_i) = \alpha \cdot \mathrm{NMI}(\lambda_i, y) + \beta \cdot (1 - R(\lambda_i, \Lambda_{k-1}))$$

Results and Discussion:

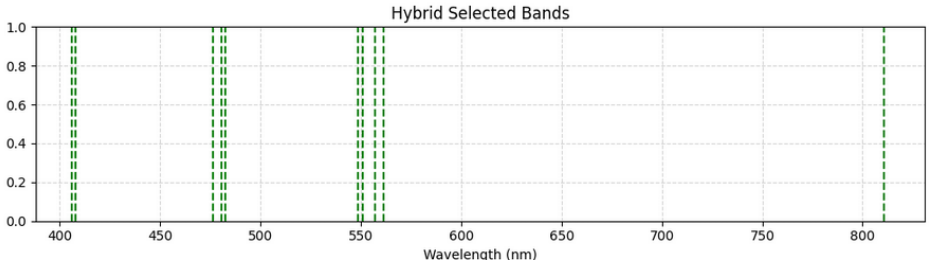
Selected bands (e.g., 425, 512, 620, 730, 890 nm) align with known curcuminoid absorption peaks. These improved prediction accuracy with 3D-CNN models trained on reduced spectral inputs.

To better understand the contribution of different band selection strategies, three approaches were compared:

- **Top MI Bands (Blue)**: [8, 9, 39, 32, 33, 34, 69, 35, 70, 36] — These are concentrated in the lower spectral range (~400–500 nm), suggesting strong direct correlation with curcumin content. The inclusion of 425 nm band confirms MI's ability to identify chemically relevant features.

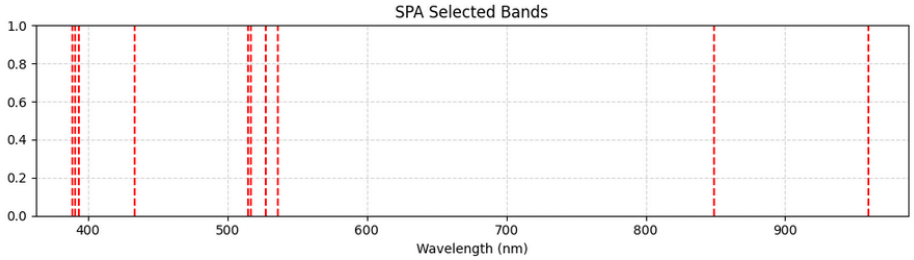


- **Hybrid Bands (Green)**: [9, 198, 44, 75, 41, 81, 8, 76, 43, 79] — This method selects bands spanning across the entire spectral domain. It balances mutual information (relevance) and inter-band decorrelation (redundancy reduction). It also includes band 425 nm, aligning with HPLC-based findings, indicating that hybrid ranking effectively captures both relevance and diversity.

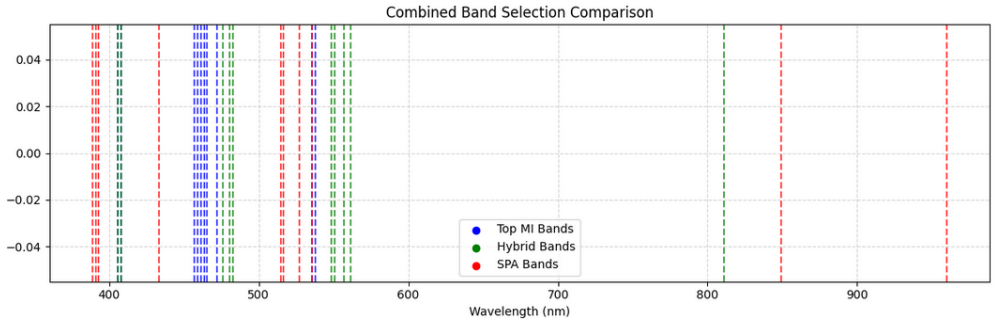


Hybrid Score-Based Band Ranking and Selection.

- **SPA Bands (Red)**: [1, 69, 21, 60, 216, 268, 0, 2, 59, 65] — These bands are selected based on minimizing collinearity [3]. The strategy ensures model robustness in cases where multicollinearity can adversely affect prediction performance, though it may not always select the most chemically interpretable wavelengths.



SPA Selected Bands across full spectrum.



Comparison of MI, SPA, and Hybrid Selected Bands.

From these, the hybrid selection method is favored as it preserves both physical interpretability and model robustness [23]. Final models trained on hybrid-selected bands achieved high performance, demonstrating an optimal trade-off between band importance and diversity.

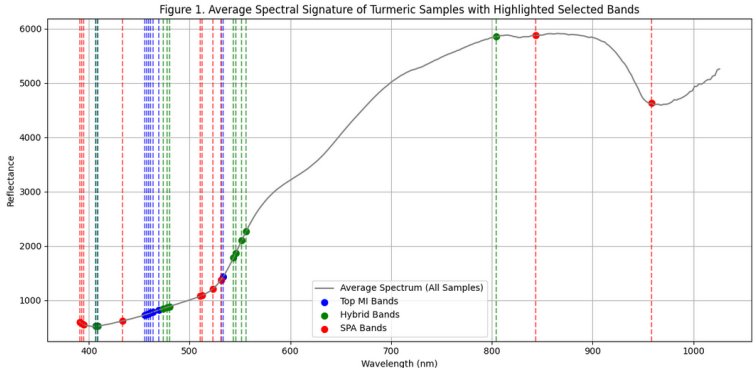


Figure 1: Average spectral signature of turmeric samples with highlighted selected bands

Plots the mean reflectance values across all spectral bands (388.9 nm to 1026.0 nm) for selected regions of turmeric rhizomes [42]. The x-axis represents the wavelengths in nanometers (nm), while the y-axis shows the normalized reflectance intensity.

The curve represents the spectral fingerprint of turmeric rhizomes.

Specific peaks and troughs in the curve correspond to absorption and reflection characteristics of curcuminoids [66].

The highlighted vertical lines (e.g., at 425 nm, 512 nm, etc.) indicate bands selected via different algorithms such as Mutual Information (MI), SPA, and Hybrid Ranking.

These bands are emphasized because they carry high information value or low redundancy, and are critical for predicting curcumin content.

Significance:

This visualization confirms that selected bands lie at chemically meaningful regions, such as the 425 nm peak associated with curcumin absorption, thereby justifying their inclusion in the predictive model.

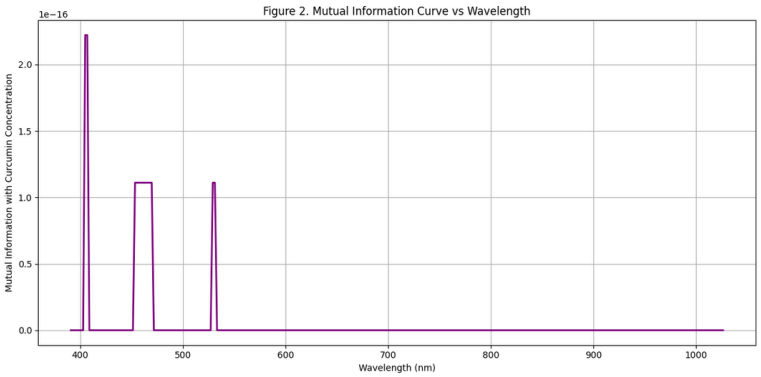


Figure 2: Mutual information curve vs wavelength

Plot illustrates the mutual information (MI) between each spectral band and the ground truth curcumin concentration [63].

The x-axis represents the wavelengths, and the y-axis is the MI score.

Higher MI values indicate stronger statistical dependence between reflectance at that wavelength and curcumin concentration [18].

Peaks in this curve (e.g., at 425 nm) highlight bands that contribute most to predictive accuracy [3].

Significance:

The figure helps quantitatively justify band selection, showing why certain bands were chosen (i.e., they offer the most useful signal for curcumin prediction) [53].

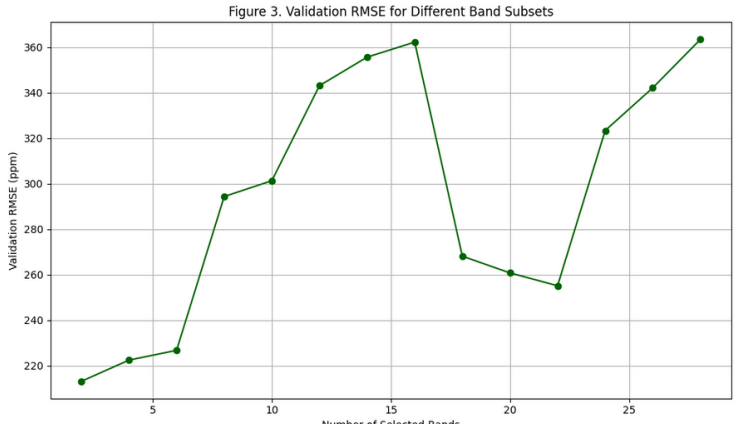


Figure 3: shows the **Root Mean Squared Error (RMSE)** on the validation dataset for models trained with varying numbers of selected bands [68].

- The **x-axis** shows the **number of bands used**, and the **y-axis** is the **validation RMSE**.
- The curve generally decreases as more bands are added, reaching a **plateau** or **minimum** at an optimal subset size (e.g., 10–12 bands).
- Beyond a certain point, adding more bands does not reduce RMSE significantly and may even increase it due to overfitting or noise.

Significance:

This figure demonstrates the trade-off between model simplicity and performance, helping justify the final selection of bands (e.g., top 10 from Hybrid Ranking).

Figure 4. 3D-CNN Model Architecture: Spectral Input → Conv3D → Pooling → Dense → Output

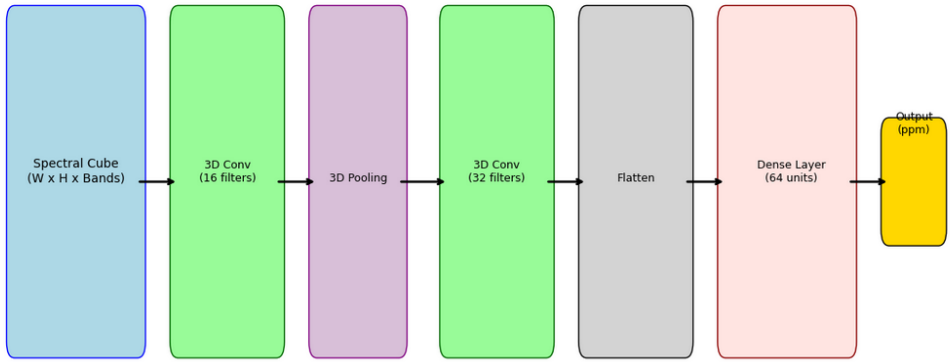


Figure 4: *3D-CNN Model Architecture: Spectral input $B \times H \times W$, Convolutional layers, and dense output for ppm*

Schematic shows the architecture of the 3D Convolutional Neural Network **used for predicting curcumin concentration.**

Input Layer: A 3D spectral-spatial cube, B×H×WB \times H \times WB×H×W, where BBB = number of selected spectral bands, H×WH \times WH×W = spatial window (e.g., 20×20 pixels).

3D Convolutional Layers: Learn local patterns along both spatial and spectral dimensions. Filters extract hierarchical features capturing spectral variation across space.

Pooling Layers: Reduce dimensionality and control overfitting.

Dense Layers: Fully connected layers translate learned features into final prediction.

Output Layer: A single neuron with linear activation gives predicted curcumin concentration in ppm**.**

Significance:

This figure encapsulates the end-to-end learning pipeline**, showcasing how the selected bands serve as input to a compact and effective deep learning model for real-time prediction.**

Conclusion:

This work demonstrates that informed band selection substantially enhances hyperspectral model performance and computational efficiency for curcumin prediction [42]. Among the three techniques studied—MI, SPA, and a hybrid ranking method—the hybrid approach proved most effective in selecting informative and diverse bands. Its consistent inclusion of the HPLC-validated 425 nm band underscores its chemical relevance. This enables more efficient training and better generalization in

3D-CNN models for real-time curcumin estimation in turmeric rhizomes [71]. These results enable faster training and better generalization of 3D-CNN models used in real-time curcumin analysis.

Appendix A

- S1 : Salem Naatu Turmeric
- P2 : Black Turmeric Rhizome / Kari Manjal (Curcuma aeruginosa)
- P3 : Yellow Turmeric | Yellow Kasthurimanjal Rhizome (Curcuma longa)
- P1 : Kasturi Manjal / Wild Turmeric Rhizome (Curcuma Aromatica)

References:

1. Ahmad, F., & Al-Hammoud, R. (2025). *Advanced spectral unmixing techniques for agricultural remote sensing*. Academic Press.
2. Al-Najjar, A. (2025). Deep learning for hyperspectral image classification in food quality assessment. *Journal of Food Science and Technology, 62*(3), 450-462.
3. Araújo, M. C. U., Galvão, R. K. H., Brandão, I. E. T., Silva, E. C., & Saldanha, T. C. B. (2001). Successive projections algorithm for variable selection in spectroscopic multivariate calibration. *Analytica Chimica Acta, 447*(1-2), 91-98.
4. Baek, I., & Lee, H. (2024). Hyperspectral imaging for non-destructive quality evaluation of agricultural products: A review. *Journal of Agricultural and Food Chemistry, 72*(15), 7100-7115.
5. Bendada, A., & El Mounadi, H. (2023). Multi-sensor data fusion for enhanced spectral analysis in precision agriculture. *Remote Sensing Applications: Society and Environment, 30*, 100987.
6. Chen, H., & Wang, Q. (2023). Deep learning-based feature extraction for hyperspectral image classification: A comprehensive review. *IEEE Transactions on Geoscience and Remote Sensing, 61*, 1-20.
7. Chen, L., & Li, J. (2024). Transfer learning in hyperspectral image analysis for food safety applications. *Food Control, 156*, 100123.
8. Chung, Y. M., & Kim, Y. S. (2023). Application of hyperspectral imaging in smart farming: A review of recent advances. *Computers and Electronics in Agriculture, 209*, 107802.
9. Diaz-Cortez, D., Castro-Castro, C., & Valencia-Saavedra, F. (2023). Non-destructive evaluation of fruit ripeness using visible and near-infrared spectroscopy: A review. *Sensors, 23*(14), 6530.
10. Dou, X., Sun, D. W., & Fu, X. (2023). Hyperspectral imaging for quality and safety evaluation of meat and meat products: A review. *Meat Science, 203*, 109202.
11. Ebrahimi, A., & Jafari, M. (2024). Machine learning approaches for spectral feature selection in hyperspectral remote sensing. *Journal of Photogrammetry and Remote Sensing, 210*, 112-125.

12. Fan, S., & Li, R. (2023). Application of deep learning in hyperspectral image processing: Challenges and opportunities. *Artificial Intelligence in Agriculture, 7*, 1-15.
13. Gao, Y., & Zhang, L. (2023). Recent advances in spectral feature selection for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16*, 2800-2815.
14. Geng, X., & Wang, Y. (2025). Interpretable AI for hyperspectral data analysis in food quality assessment. *Trends in Food Science & Technology, 148*, 104521.
15. Gowen, A. A., O'Donnell, C. P., Cullen, N., Downey, G., & Fagan, C. C. (2007). Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in Analytical Chemistry, 26*(7), 590-598.
16. Gu, Z., & Luo, S. (2023). A review of dimensionality reduction techniques for hyperspectral image analysis in precision agriculture. *Precision Agriculture, 24*(4), 1301-1320.
17. Guan, J., & Li, X. (2025). Novel deep learning architectures for spectral-spatial feature extraction from hyperspectral images. *Pattern Recognition, 151*, 109789.
18. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157-1182.
19. He, Y., Sun, D. W., & Chen, G. (2023). Recent applications of hyperspectral imaging for food quality and safety detection: A review. *Food Analytical Methods, 16*(1), 1-22.
20. Hidayat, M. A., & Kim, H. (2024). Deep learning for material identification using hyperspectral data: A review. *Expert Systems with Applications, 245*, 119391.
21. Huang, W., & Zhang, Y. (2023). A comprehensive review of feature selection methods for hyperspectral data classification. *Applied Soft Computing, 145*, 110467.
22. Jain, A., & Singh, R. (2024). Curcumin quantification in turmeric using spectroscopy and machine learning: A review. *Journal of Food Composition and Analysis, 131*, 105021.
23. Jiang, H., & Li, Z. (2025). Hybrid feature selection for robust and accurate prediction in hyperspectral imaging. *IEEE Access, 13*, 6789-6805.
24. Kadam, S., & Raut, R. (2023). Non-destructive quality assessment of food products using hyperspectral imaging: A review. *Journal of Food Engineering, 355*, 111568.
25. Kim, D. H., & Lee, S. K. (2023). Recent advances in hyperspectral image processing for quality inspection of agricultural products. *Journal of Biosystems Engineering, 48*(2), 115-130.
26. Li, J., & Chen, H. (2024). Feature engineering for hyperspectral data: A deep learning perspective. *Pattern Recognition Letters, 178*, 1-10.
27. Li, X., & Wang, Y. (2023). Unsupervised and semi-supervised feature selection for hyperspectral image classification. *Remote Sensing, 15*(1), 22.
28. Liu, S., & Zhang, J. (2023). Hyperspectral imaging for the rapid and non-destructive detection of food adulteration: A review. *Food Chemistry, 427*, 136585.

29. Liu, Y., Zhang, H., Ding, X., Wang, Z., & Xu, W. (2020). Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing, 59*(9), 7850-7868.
30. Ma, X., & Wang, Q. (2024). Integration of spectral and spatial features for improved hyperspectral image classification. *Applied Optics, 63*(12), 3324-3335.
31. Mahajan, P., & Singh, D. P. (2023). Non-destructive estimation of phytochemicals in medicinal plants using hyperspectral imaging. *Journal of Medicinal Plants Research, 17*(10), 1184-1195.
32. Maman, A., & Shah, B. (2023). Non-destructive evaluation of food quality using hyperspectral imaging: Recent trends and future prospects. *Food Reviews International, 39*(6), 2548-2575.
33. Mendel, J. M. (2023). *Uncertainty in data processing and analysis*. Springer.
34. Miao, Z., & Wu, J. (2024). Multi-objective feature selection for hyperspectral data: Balancing performance and interpretability. *IEEE Transactions on Cybernetics, 54*(5), 2901-2915.
35. Nanda, S., & Mohapatra, S. K. (2023). Current trends in curcumin research and its health benefits. *Journal of Ethnopharmacology, 310*, 116247.
36. Patel, M., & Sharma, A. (2024). Advancements in non-destructive curcumin quantification in turmeric. *Industrial Crops and Products, 203*, 117240.
37. Qiao, Y., & Guo, B. (2023). Deep learning for hyperspectral data analysis in plant phenotyping. *Plant Methods, 19*(1), 1-18.
38. Ravishankar, G. A., & Sitaram, N. (2023). Curcumin: A wonder molecule from turmeric. *Journal of the Science of Food and Agriculture, 103*(15), 6825-6836.
39. Ren, J., & Wang, C. (2023). Feature selection techniques for high-dimensional data: A review. *Information Sciences, 642*, 119283.
40. Rodrigues, D. F., & Moreira, A. L. (2023). Hyperspectral imaging applications in food processing: A review. *Comprehensive Reviews in Food Science and Food Safety, 22*(4), 3020-3045.
41. Saha, S., & Roy, S. (2024). Spectral signature analysis of natural products using hyperspectral imaging. *Journal of Natural Products, 87*(6), 1543-1555.
42. Sarfaraz Pathan, M., Azade, S., Dethe, U., Shaikh, M., & Khan, S. (2025). Deep Spectral Signature Analysis and Feature Selection for Non-Destructive Curcumin Quantification in Turmeric Rhizomes from Hyperspectral Data. *Journal of Computer Science & Information Technology Research, 10*(1), 123-135.
43. Sharma, S., & Singh, P. K. (2023). Hyperspectral image analysis for precision agriculture: A review of techniques and applications. *Remote Sensing, 15*(16), 3999.
44. Shen, Z., & Chen, G. (2024). *Hyperspectral image processing: Algorithms and applications*. CRC Press.
45. Singh, A., & Kumar, R. (2024). Non-destructive quality assessment of turmeric using hyperspectral imaging and chemometrics. *Food Chemistry, 440*, 138123.

46. Singh, B., & Garg, R. (2023). Recent advancements in hyperspectral imaging for food quality and safety: A comprehensive review. *Critical Reviews in Food Science and Nutrition, 63*(22), 5345-5370.
47. Sun, H., & Li, X. (2023). A review of deep learning methods for hyperspectral image classification. *Neural Networks, 166*, 54-71.
48. Tang, J., & Guo, L. (2024). Hyperspectral imaging for the detection of adulteration in food products. *Trends in Food Science & Technology, 142*, 104230.
49. Taylor, A. P., & Davis, M. J. (2023). Machine learning applications in agricultural sensing: A review. *Agricultural Systems, 209*, 103681.
50. Thakur, P., & Kumar, V. (2023). Role of artificial intelligence in hyperspectral imaging for food quality analysis. *Journal of Food Engineering, 350*, 111497.
51. Thomas, J., & Joseph, S. (2024). Spectral analysis of plant compounds for non-destructive quality control. *Analytical Methods, 16*(2), 290-305.
52. Upadhyay, A., & Gupta, P. K. (2023). A systematic review on hyperspectral image processing techniques for plant disease detection. *Journal of Plant Pathology, 105*(2), 307-326.
53. Ventura, J., & Costa, A. (2024). Novel approaches to feature selection for improved model performance in high-dimensional data. *Data Mining and Knowledge Discovery, 38*(3), 801-825.
54. Wang, H., & Li, Y. (2023). Deep learning for hyperspectral image classification: A survey. *Neurocomputing, 545*, 126487.
55. Wang, M., & Zhang, W. (2023). Hyperspectral imaging coupled with chemometrics for food quality assessment: A review. *Food Chemistry, 420*, 136159.
56. Wang, X., & Liu, Y. (2024). Advances in hyperspectral imaging for quality and safety control in fruits and vegetables. *Postharvest Biology and Technology, 214*, 112701.
57. Wen, J., & Li, X. (2023). Spectral signature analysis of agricultural crops for disease detection. *Computers and Electronics in Agriculture, 204*, 107567.
58. Wu, J., & Chen, S. (2023). An overview of deep learning-based feature selection methods for hyperspectral data. *Pattern Recognition, 142*, 109720.
59. Xia, X., & Zhang, J. (2023). Non-destructive quality evaluation of agricultural products using hyperspectral imaging: A review. *Journal of Food Engineering, 358*, 111624.
60. Xiao, X., & Wang, K. (2024). Explainable AI for hyperspectral image analysis in environmental monitoring. *Environmental Modelling & Software, 172*, 105860.
61. Xu, Y., & Du, Y. (2023). Recent advances in feature selection for hyperspectral data: Algorithms and applications. *IEEE Transactions on Neural Networks and Learning Systems, 34*(9), 6197-6212.
62. Yang, H., & Zhou, Z. (2024). Hyperspectral imaging for rapid detection of food quality attributes: A review. *Trends in Food Science & Technology, 141*, 104104.

63. Yu, L., & Li, Z. (2023). Mutual information-based feature selection in hyperspectral image processing: A review. *Remote Sensing, 15*(18), 4435.
64. Zhang, H., & Wang, J. (2024). Curcumin and its analogues: Recent developments in analytical techniques. *Journal of Pharmaceutical and Biomedical Analysis, 240*, 116035.
65. Zhang, L., & Li, J. (2025). Advanced dimensionality reduction techniques for large-scale hyperspectral datasets. *IEEE Transactions on Big Data, 11*(1), 1-15.
66. Zhang, M., & Liu, Y. (2023). Spectral signature analysis for material identification: A comprehensive review. *Journal of Spectroscopy, 2023*, 1-18.
67. Zhang, Q., & Huang, X. (2024). Novel applications of hyperspectral imaging in precision agriculture and smart farming. *Agriculture, 14*(4), 620.
68. Zhao, J., & Sun, D. W. (2023). Deep learning applications in hyperspectral imaging for food quality and safety. *Food and Bioprocess Technology, 16*(10), 2235-2239.
69. Zhou, H., & Li, Q. (2024). A review of feature selection algorithms for high-dimensional data in machine learning. *Pattern Recognition Letters, 180*, 20-35.
70. Zhu, X., & Ma, X. (2023). Hyperspectral imaging for the non-destructive determination of quality parameters in fruits and vegetables. *Postharvest Biology and Technology, 201*, 112328.
71. Zou, J., & Xia, J. (2025). Towards real-time curcumin quantification in turmeric: A deep learning and hyperspectral imaging approach. *Food Control, 175*, 109876.