GRADIENT-BASED BAND SELECTION FOR HYPERSPECTRAL IMAGING WITH NEURAL NETWORKS

Nicolas Termote[1,2]*, Aravind Krishnaswamy Rangarajan*[2]*, Naftali Slob*[2]*, Bas Boom*[2]

[1]Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
[2]OnePlanet Research Center, Plus Ultra II, Bronland 10, 6708 WE Wageningen, The Netherlands
aravind.krishnaswamyrangarajan@imec.nl

ABSTRACT

Hyperspectral imaging (HSI) provides rich spectral information for applications such as precision agriculture, mineral exploration, and environmental monitoring. However, the high dimensionality and cost of hyperspectral systems have driven interest in band selection for efficient multispectral imaging. This study introduces IRB-BECI, an iterative backpropagation-driven band selection method integrated with U-Net for hyperspectral image classification. IRB-BECI improves classification accuracy while reducing spectral dimensionality and consistently outperforms PCA Loadings and FB-BECI under reduced-band conditions. Experiments on four benchmark datasets show gains of 17.5% for LIB-HSI Facades, 16.3% for Pavia University, and 10.2% for Indian Pines, while performance on Salinas remains comparable to PCA-based methods. These results highlight the effectiveness of gradient-based selection in preserving classification performance under severe compression, offering a scalable and cost-effective solution for hyperspectral imaging.

Index Terms— Neural Networks, Hyperspectral Imaging (HSI), Saliency, Feature extraction, Band selection, Gradient Backpropagation

1. INTRODUCTION

Hyperspectral imaging (HSI) has emerged as a transformative technology, capturing rich spectral information across numerous wavelength bands and assembling them into 3D hypercubes that encapsulate both spatial and spectral characteristics of objects. This capability has enabled breakthroughs in fields such as astronomy, medical diagnosis, agriculture, environmental science, and mineral detection, where precise pixel-level classification is essential[1].

Deep learning, particularly convolutional neural networks (CNNs), has significantly advanced HSI classification by enabling joint spectral–spatial feature learning. Early approaches using 2D CNNs were limited in their ability to fully exploit spectral information, prompting the development of 3D CNNs and hybrid models that more effectively integrate spatial and spectral cues [2]. Despite these advances, HSI classification remains challenging because of the high dimensionality of spectral data and limited training samples, leading to overfitting and increased computational demands. Redundant information in adjacent bands further exacerbates data volume and processing complexity [3].

A broad spectrum of approaches has been proposed. Linear projection techniques such as Principal Component Analysis (PCA) provide efficient compression but discard physical interpretability of spectral bands [4]. Nonlinear manifold learning methods aim to preserve intrinsic data geometry but are not inherently task-driven [5]. Information-theoretic and wrapper-based methods explicitly target discriminative bands but often rely on handcrafted criteria and miss higher-order spectral interactions [6, 7].

Deep learning-based approaches address these limitations by learning nonlinear representations directly from data. Methods based on autoencoders and related models focus on latent feature learning rather than explicit band selection [8]. More recent CNN-driven strategies incorporate band selection via gradient-based attribution such as saliency maps [9, 10]. Reinforcement learning approaches enable adaptive selection but introduce higher computational cost and sensitivity to reward design [11]. Overall, existing methods either decouple band selection from the classifier or insufficiently exploit spectral–spatial interactions.

We propose gradient-based band selection methods that integrate channel ranking into CNN training via backpropagation. By iteratively refining spectral subsets (Fig. 1), the approach yields compact, discriminative bands and improves reduced-band classification, particularly in low-band regimes, while remaining compatible with spectral–spatial architectures [12].

2. METHODOLOGY

In this study, the classification system based on a U-Net architecture remains unchanged across all experiments. The main methodological difference lies in the approach to band selection: PCA Loadings uses principal component analysis to rank channel importance, while Full Band Backprop-

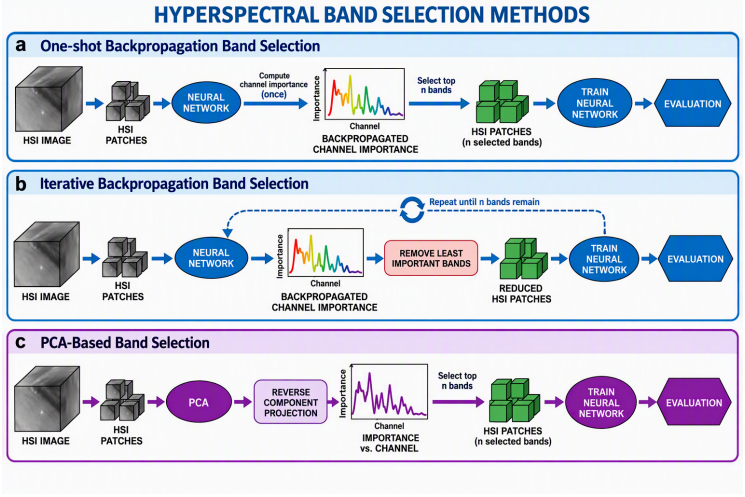


Fig. 1. Schematic representation of the methodology and experimental workflow: (a) Full Band Backpropagation Enhanced Channel Identification (FB-BECI), (b) Iterative Reduced Band Backpropagation Enhanced Channel Identification (IRB-BECI), and (c) PCA Loadings (PCAL).

agation Enhanced Channel Identification (FB-BECI) and Iterative Reduced Band Backpropagation Enhanced Channel Identification (IRB-BECI) employ gradient-based relevance maps from network backpropagation. IRB-BECI further refines band selection through iterative reduction and retraining. Hyperspectral images are divided into patches, and reduced-band datasets are created using the most informative channels identified by each method. The U-Net classifier is then trained on these datasets, allowing a direct comparison of classification results. This design isolates the impact of band selection strategy, ensuring that differences in performance are attributable solely to the patch and band selection methods.

2.1. Datasets & Preprocessing

In this study, four hyperspectral datasets are used: **Pavia University (PU)** [13], **Indian Pines** [14], **Salinas Scene** [15], and **LIB-HSI façades** [16]. All datasets contain a limited number of images and exhibit varying spatial resolutions, making patch-based sampling essential to increase training samples, preserve spatial context, and ensure uniform input sizes under practical memory constraints.

For PU, Indian Pines, and Salinas, 10,000 random patches of size 16×16 with random rotations are generated to enhance orientation invariance and reduce sampling bias [17]. For LIB-HSI, each image is split into 10 random patches of size 32×32 after applying an 80–20 train/test split. All datasets follow the same 80–20 split strategy.

To evaluate the impact of band selection, reduced-band datasets are created using six compression ratios: 2, 10, 20, 40, 80, and 160. These subsets allow systematic analysis of classification performance under varying spectral dimensionality. The implemented band selection strategies are detailed in Section 2.4.

2.2. Network and Training Configuration

The baseline model used in this study is a 4-level U-Net with convolutional filters $\{32, 64, 128, 256\}$ and a 512-filter bottleneck. Each stage uses ReLU activations, 2×2 max-pooling for downsampling, and transposed convolutions for upsampling; the output layer applies sigmoid activation. Training uses cross-entropy loss with class weights (Eq. 2):

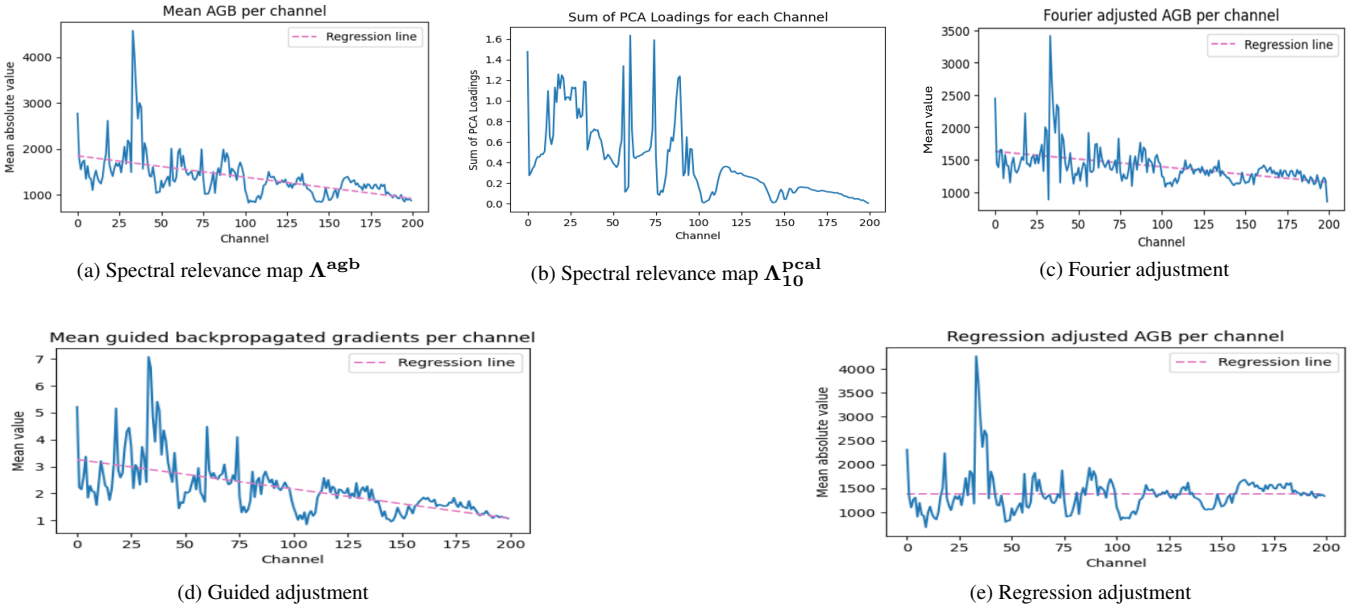
$$\omega_c = \frac{\sum_{i=1}^{N_{px}} \mu_c(i)}{N_{px}} \qquad \mu_c(i) = \begin{cases} 1 & \text{if } g(i) = c \\ 0 & \text{if } g(i) \neq c \end{cases} \tag{1}$$

$$L = -\sum_{c=1}^{C} \omega_c \sum_{i=1}^{N_{px}} y_{i,c} \log(p_{i,c}) \tag{2}$$

With N_{px} being the number of pixels, ω_c being the weight of class c, and $g(i)$ being the class of pixel i. $y_{i,c}$ is one-hot ground-truth indicator for pixel i and class c ($y_{i,c} = 1$ if pixel i belongs to class c, otherwise 0). $p_{i,c}$ is predicted softmax probability of pixel i belonging to class c.

Optimization: SGD with learning rate 0.01 (Facades) and 0.1 (others), momentum 0.9, batch size 64, weight decay 5×10^{-4}. U-Net is trained for 150 epochs (100 for LIB-HSI). To ensure reproducibility, random seed of 42 is fixed. For LIB-HSI, eight metaclasses group of similar materials to simplify segmentation without affecting comparative analysis.

Each experiment runs five times; results from the last 10 epochs are averaged after convergence. Data augmentation (random rotations and re-patching per epoch) mitigates overfitting. All experiments use PyTorch 2.2.1 on a system with 32 GB RAM and NVIDIA RTX A3000 GPU (6 GB VRAM).



(a) Spectral relevance map $\mathbf{\Lambda^{agb}}$

(b) Spectral relevance map $\mathbf{\Lambda^{pcal}_{10}}$

(c) Fourier adjustment

(d) Guided adjustment

(e) Regression adjustment

Fig. 2. Spectral relevance maps and adjustment methods for Indian Pines.

Algorithm 1 IRB-BECI: Iterative Reduction-Based Band Selection using Absolute Gradient Backpropagation

$\mathbf{X} \leftarrow$ full-band hyperspectral input ▶ input data
$B \leftarrow$ initial number of spectral bands ▶ band count
$\mathbf{r} \leftarrow$ compression ratios ▶ reduction factors
$j \leftarrow 0$ ▶ iteration index j
$\mathcal{S} \leftarrow \emptyset$ ▶ selected band sets $\mathcal{S}$
repeat
  Train network $\mathbf{N}(\mathbf{X})$ ▶ model training
  $\mathbf{\Lambda} \leftarrow |\nabla_{\mathbf{X}} \mathcal{L}|$ ▶ $\nabla_{\mathbf{X}}$: grad w.r.t. $\mathbf{X}$; $\mathcal{L}$: loss
  $\mathbf{I} \leftarrow \text{argsort}(\mathbf{\Lambda})$ ▶ band indices I (ranked)
  $B' \leftarrow \text{round}(B/r_j)$ ▶ reduced bands B'
  $\mathbf{I}' \leftarrow \mathbf{I}(1{:}B')$ ▶ top-B' bands
  Append $\mathbf{I}'$ to $\mathcal{S}$ ▶ update $\mathcal{S}$
  $\mathbf{X} \leftarrow \mathbf{X}(:, \mathbf{I}')$ ▶ channel reduction
  $B \leftarrow B', j \leftarrow j+1$ ▶ update B, j
until $B = 1$
return $\mathcal{S}$ ▶ final band selections

Performance is evaluated using classification accuracy and the pixel-wise F1 score computed per class (Eq. 3):

$$F1 = \frac{\text{True Positives}}{\text{True Positives} + \frac{1}{2}(\text{False Positives} + \text{False Negatives})} \quad (3)$$

2.3. Band Selection Technique

2.3.1. Strategy I - PCA Loadings (PCAL)

PCA mixes information across all channels, making direct band selection non-trivial. To retain interpretability, a PCA-based loading strategy is adopted [19]. After computing PCA, the loadings matrix $\mathbf{V}$ represents each principal component as a linear combination of original bands. The importance of band b is estimated by aggregating its absolute contributions across the top k components:

$$\Lambda^{\text{pcal}}(b) = \sum_{p=1}^{k} |V_{b,p}| \quad (4)$$

Bands are ranked based on $\Lambda^{\text{pcal}}(b)$.

2.3.2. Strategy II - Full Band Backpropagation Enhanced Channel Identification (FB-BECI)

Gradient backpropagation enables estimation of input feature relevance by propagating the loss gradient to the input layer, thereby quantifying the sensitivity of the model output to each spectral band. Let the input hyperspectral image be $\mathbf{X} \in \mathbb{R}^{m \times n \times s}$, corresponding to the input layer ($\mathbf{z}^1$) of the network. For an input tensor $\mathbf{z}^1 \in \mathbb{R}^{m \times n \times s}$, where m, n denote spatial dimensions and s the number of spectral bands, the gradient of the loss C with respect to the input captures the contribution of each pixel-band location.

To obtain a global band importance measure, we define Absolute Gradient Backpropagation (AGB) as:

$$\Lambda_N^{\text{AGB}}(b) = \frac{1}{nm} \sum_{i,j=1}^{m,n} \left| \frac{\partial C}{\partial z^1_{i,j,b}} \right| \quad (5)$$

where $\frac{\partial C}{\partial z^1_{i,j,b}}$ represents the gradient of the loss with respect to the input at spatial location (i, j) and spectral band b. The ab-

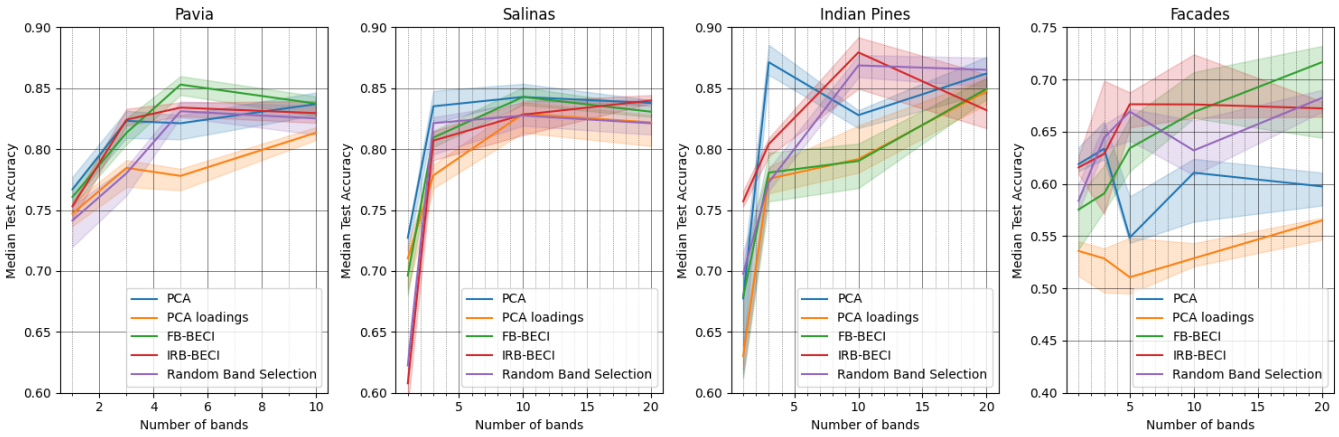


Fig. 3. Median pixel accuracy score across all classes, reduced by different Band Selection Methods. The results are illustrated with curves that include a 95% confidence interval, determined via bootstrapping [18].

solute operation ensures that both positive and negative contributions are accounted for, while spatial averaging produces a band-wise relevance score independent of location.

While AGB provides a robust and intuitive measure of band relevance, raw gradient-based scores may still be affected by systematic trends arising from sensor characteristics, wavelength ordering, or network-induced biases. To enhance robustness and interpretability, several complementary refinement strategies are explored.

Guided Backpropagation (GBP): Guided backpropagation modifies standard backpropagation by introducing a double-gating mechanism at ReLU layers, such that only positive gradients from positive activations are propagated:

$$\left.\frac{\partial C}{\partial z_j^l}\right|_{\text{GBP}} = \mathbb{1}\{a_j^l > 0\}\,\mathbb{1}\left\{\frac{\partial C}{\partial a_j^l} > 0\right\}\frac{\partial C}{\partial a_j^l} \tag{6}$$

where $a_j^l = \max(0, z_j^l)$. The corresponding band-importance score is computed as

$$\Lambda_N^{\text{GBP}}(b) = \frac{1}{n \times m}\sum_{i,j=1}^{m,n}\left|\left.\frac{\partial C}{\partial z_{i,j,b}^1}\right|_{\text{gbp}}\right| \tag{7}$$

GBP typically produces sharper saliency maps by emphasizing positively contributing features, but it may suppress informative negative contributions that are retained by AGB.

Regression-Adjusted AGB: To mitigate linear wavelength-dependent trends that can bias $\Lambda_N^{\text{AGB}}(b)$, a first-order polynomial $f(b) = \alpha b + \beta$ is fitted to the importance curve and subtracted:

$$\Lambda_{\text{adj}}(b) = \Lambda_N^{\text{AGB}}(b) - f(b) + \overline{f} \tag{8}$$

where $\overline{f}$ is the mean of $f(b)$ over all bands. This normalization preserves the global scale of the importance scores while removing linear drifts associated with spectral ordering (see Figure.2.

Fourier-Adjusted AGB: Alternatively, low-frequency components in the importance vector $\mathbf{\Lambda}$ can be suppressed using frequency-domain filtering:

$$\mathbf{F} = \text{FFT}(\mathbf{\Lambda}), \quad F_k \leftarrow 0 \;\text{ for }\; k \leq B/4 \tag{9}$$

$$\Lambda_{\text{fourier}}(b) = \Re(\text{IFFT}(\mathbf{F}))_b \tag{10}$$

where B is the total number of spectral bands and F_k denotes the k-th Fourier coefficient of the band importance vector. This operation removes slowly varying trends while preserving higher-frequency variations that may reflect class-specific spectral interactions.

2.3.3. Strategy III - Iterative Reduction-Based Band Selection using Absolute Gradient BackpropagationIRB-BECI

To further refine selection, an iterative reduction strategy is employed. Starting from the full spectrum, bands are ranked using gradient scores, and only the top subset is retained. The model is retrained on the reduced input, and the process is repeated until a target dimensionality is reached. This progressively refines band importance under reduced spectral context, yielding compact and discriminative subsets. The pseudo-code for this method can be seen in Algorithm 1.

3. RESULTS

This section presents a comprehensive evaluation of the proposed band-selection strategies on four benchmark hyperspectral datasets, focusing on reduced-band classification performance. Results are reported as median pixel-wise F1 scores over repeated runs, computed on foreground classes to avoid background bias.

Figure 2 compares the spectral relevance maps generated by raw AGB, PCAL, and AGB refinement techniques for the Indian Pines dataset. While regression and Fourier adjustments smooth or detrend the curves and guided backpropagation sharpens peaks, the dominant informative spectral regions remain consistent. Quantitatively, regression improves performance at 20 and 5 bands, whereas raw AGB performs best at 10 and 3 bands. As no refinement consistently outperforms AGB, all subsequent experiments use raw AGB.

All models were trained using identical settings and evaluated over multiple runs. FB-BECI was more computationally efficient than IRB-BECI (6,225 s vs. 9,450 s on Indian Pines and 31,050 s vs. 46,650 s on LIB-HSI Facades), reflecting the overhead of iterative retraining and channel-importance recomputation. The higher runtime on LIB-HSI Facades is attributed to its larger size and multi-file storage structure despite the use of pinned memory, while inference remained efficient at only 1.5-3 ms per sample.

Figure 3 shows accuracy as a function of band count. Across datasets, performance improves rapidly from very low band counts and saturates beyond 10–20 bands. Indian Pines is more sensitive to reduction, while LIB-HSI Facades achieves lower overall accuracy but follows similar trends. In the low-band regime (< 10 bands), FB-BECI and IRB-BECI consistently outperform PCAL and random selection, with diminishing differences as band count increases.

Table 1 summarizes the main quantitative results. Gradient-based methods outperform PCAL in most reduced-band settings. IRB-BECI achieves the best performance on Indian Pines (0.874, 0.883, 0.821, 0.724 at 20, 10, 3, and 1 bands) and under strong compression on Pavia and Facades. FB-BECI performs best on Salinas at moderate band counts and on Facades at higher band budgets. PCAL remains competitive primarily in extreme one-band scenarios. Statistical tests confirm significant differences across most configurations.

Relative to PCAL, FB-BECI improves F1 scores by 1.2% (Salinas), 17.1% (Facades), 4.9% (Indian Pines), and 13.4% (Pavia). IRB-BECI achieves comparable or larger gains, particularly under strong compression, highlighting the benefit of iterative refinement.

At 10 bands, IRB-BECI selects dataset-specific subsets that align with prior studies: overlap is observed with reported informative bands in Salinas and mid-spectrum regions in Pavia, while Indian Pines shows partial agreement due to variability in preprocessing and selection criteria [20, 21, 22, 23].

4. CONCLUSION

IRB-BECI, an iterative backpropagation-driven band selection method, demonstrates superior performance under reduced-band conditions across multiple hyperspectral datasets, with notable gains on LIB-HSI Facades, Pavia University, and Indian Pines, while remaining comparable to PCAL on Salinas. The method is particularly effective in low-band scenarios, offering improved robustness at the cost of increased computational time. Overall, gradient-based channel selection integrated with model learning enables substantial dimensionality reduction without sacrificing classification performance in U-Net.

Table 1. Median test F1 scores across all datasets for different channel configurations and methods.

Pavia University

Ch.	PCAL	FB-BECI	IRB-BECI
103	*0.563*	*0.563*	*0.563*
50	0.523$^{\Downarrow}$	**0.679**$^{\Uparrow}$	**0.679**$^{\Uparrow}*$
10	0.493$^{\Downarrow}$	**0.615**$^{\Uparrow}$	0.586$^{\Uparrow}*$
3	0.568	0.532	**0.585**$^{\Uparrow}*$
1	**0.482**	0.448	0.453

Salinas

Ch.	PCAL	FB-BECI	IRB-BECI
224	*0.869*	*0.869*	*0.869*
20	0.835$^{\Downarrow}$	0.845$^{\uparrow}$	**0.879**$^{\Uparrow}*$
10	0.838	**0.861**	0.847
3	0.781$^{\Downarrow}$	**0.849**$^{\Uparrow}$	0.812$^{\Uparrow}*$
1	**0.675**$^{\Uparrow}$	0.627$^{\Downarrow}$	0.602$^{\Downarrow}*$

Indian Pines

Ch.	PCAL	FB-BECI	IRB-BECI
200	*0.888*	*0.888*	*0.888*
20	0.862	0.862	**0.874**$^{\uparrow}*$
10	0.847	0.851	**0.883**$^{\Uparrow}*$
3	0.752$^{\Downarrow}$	0.796$^{\Uparrow}$	**0.821**$^{\Uparrow}*$
1	0.575$^{\Downarrow}$	0.651$^{\Uparrow}$	**0.724**$^{\Uparrow}*$

LIB-HSI Facades

Ch.	PCAL	FB-BECI	IRB-BECI
204	*0.658*	*0.658*	*0.658*
20	0.543$^{\Downarrow}$	**0.672**$^{\Uparrow}$	0.638$^{\Uparrow}*$
10	0.531$^{\Downarrow}$	**0.654**$^{\Uparrow}$	0.629$^{\Uparrow}*$
3	0.495$^{\Downarrow}$	0.550$^{\uparrow}$	**0.579**$^{\Uparrow}*$
1	0.530$^{\downarrow}$	0.583$^{\uparrow}$	**0.584**$^{\Uparrow}*$

Non-parametric statistical analysis is conducted due to non-normal data distributions. Overall differences are evaluated using the Kruskal–Wallis test, followed by Dunn's and Conover's post-hoc tests with Benjamini–Hochberg correction [24, 25, 26]. Asterisks denote $p < 0.01$, while arrows indicate the direction and significance of pairwise changes: $\rightarrow$ ($p < 0.1$), $\Rightarrow$ ($p < 0.05$), $\Rrightarrow$ ($p < 0.01$).

5. ACKNOWLEDGEMENT

This research was supported with funding from the Dutch Growth Fund Nxtgen Hightech Agrifood Use Case 7 Hands-free Orchards

6. REFERENCES

[1] Arun A. Solomon and Akila S. Agnes, "Land-cover classification with hyperspectral remote sensing image using cnn and spectral band selection," *Remote Sensing Applications: Society and Environment*, vol. 31, pp. 100986, 2023.

[2] Nooshin Noshiri, Michael A. Beck, Christopher P. Bidinosti, and Christopher J. Henry, "A comprehensive review of 3d convolutional neural network-based classification techniques of diseased and defective crops using non-uav-based hyperspectral images," *Smart Agricultural Technology*, vol. 5, pp. 100316, 2023.

[3] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE geoscience and*

remote sensing magazine, vol. 7, no. 2, pp. 159–173, June 2019.

[4] Pedram Ghamisi, Naoto Yokoya, Jun Li, Wenzhi Liao, Sicong Liu, Javier Plaza, Behnood Rasti, and Antonio Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, pp. 37–78, Dec. 2017.

[5] Sam T. Roweis and Lawrence K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[6] Zhong-Min Wang, Shu-Yuan Hu, and Hui Song, "Channel selection method for eeg emotion recognition using normalized mutual information," *IEEE Access*, vol. 7, pp. 143303–143311, 2019.

[7] Chris Ding and Hanchuan Peng, "Minimum redundancy feature selection from microarray gene expression data," 09 2003, vol. 3, pp. 523– 528.

[8] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[9] Lei Zhao, Yi Zeng, Peng Liu, and Xiaohui Su, "Band selection with the explanatory gradient saliency maps of convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 12, pp. 2105–2109, 2020.

[10] Lei Zhao, Yi Zeng, Peng Liu, and Guojin He, "Band selection via explanations from convolutional neural networks," *IEEE Access*, vol. 8, pp. 56000–56014, 2020.

[11] Lichao Mou, Sudipan Saha, Yuansheng Hua, Francesca Bovolo, Lorenzo Bruzzone, and Xiao Xiang Zhu, "Deep reinforcement learning for band selection in hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[12] Giorgio Morales, John Sheppard, Riley Logan, and Joseph Shaw, "Hyperspectral band selection for multispectral image classification with convolutional networks," in *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1–8.

[13] Muhammad Ahmad, Stanislav Protasov, Adil Mehmood Khan, Rasheed Hussain, Asad Masood Khattak, and Wajahat Ali Khan, "Pavia university dataset," Dataset, PLOS ONE Supplementary Material, Jan. 2018.

[14] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe, "220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3," Sept. 2015.

[15] "Earth Observation, salinas hyperspectral dataset," .

[16] Nariman Habili, Ernest Kwan, Weihao Li, Jeremy Oorloff, Christfried Webbers, Ali Armin, and Lars Petersson, "A hyperspectral and rgb dataset for building façade segmentation," in *ECCV Workshop on Computer Vision for Civil and Infrastructure Engineering*, 2022.

[17] Wei Li, Chen Chen, Mengmeng Zhang, Hengchao Li, and Qian Du, "Data augmentation for hyperspectral image classification with deep cnn," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 4, pp. 593–597, 2019.

[18] Thomas J. DiCiccio and Bradley Efron, "Bootstrap confidence intervals," *Statistical Science*, vol. 11, no. 3, pp. 189–212, 1996.

[19] Chein-I Chang, Qian Du, Tzu-Lung Sun, and M.L.G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2631–2641, 1999.

[20] Giorgio Morales, John W. Sheppard, Riley D. Logan, and Joseph A. Shaw, "Hyperspectral dimensionality reduction based on inter-band redundancy analysis and greedy spectral selection," *Remote Sensing*, vol. 13, no. 18, pp. 3649, 2021.

[21] T. Hu et al., "Hyperspectral band selection method based on global partition clustering," *Remote Sensing*, vol. 17, no. 3, pp. 435, 2025.

[22] Xin Wang and Guoqiang Wang, "Hyperspectral band selection based on adaptive neighborhood grouping and local structure correlation," *Journal of Sensors*, p. 5530385, 2021.

[23] Nariman Habili, Ernest Kwan, Weihao Li, Christfried Webers, Jeremy Oorloff, Mohammad Ali Armin, and Lars Petersson, "A hyperspectral and rgb dataset for building façade segmentation," in *Computer Vision – ECCV 2022 Workshops*, 2023.

[24] W J Conover and R L Iman, "Multiple-comparisons procedures. informal report," *Journal Name*, Feb. 1979.

[25] Olive Jean Dunn, "Multiple comparisons using rank sums," *Technometrics*, vol. 6, no. 3, pp. 241–252, 1964.

[26] Yoav Benjamini and Yosef Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.