Robust Hyperspectral Crop Classification Using Stable Band-Pair Learning Across Field Conditions

S. Lavanya[1], *, M. Sutharsan[2]
[1]Assistant Professor/Dept. of Biomedical Engineering, Jerusalem College of Engineering, Chennai – 600100,
[2]Assistant Professor, ECE Dept, Dhirajlal Gandhi College of Technology, Salem-636309
Email: lavanyas.bme@jerusalemengg.ac.in

Abstract-Adjacent training and test pixels can have almost the same spectra and be split into different subsets in a random fashion, resulting in optimistic estimates of reliability for hyperspectral crop classification. This paper introduces a two-band extension of the Indian Pines logistic baseline called field-stable band-pair learning (FS-BPL). FS-BPL clusters the same units (crop fields), normalizes the between-field distance for all bands by the distance within fields, and identifies pairs of bands with low redundancy via nested complete-field validation. FS-BPL had a mean balanced accuracy of 68.4% across 12 holdout pasture-tree fields, 10.2 percentage points better than the mean baseline fixed two-band accuracy. This result also has to be understood in the context of the 88.6% superpixel-based selector accuracy reported for uniformly distributed pixel sampling across 16 target classes by an evaluated 2025 superpixel-based selector. The proposed solution yields an easily disclosed classifier with minimal sensors and a field-level estimate of generalization.

Keywords: Hyperspectral imaging, classification, band selection, field holdout, spatial and spectral, logistic regression, interpretability.

I. Introduction

A hyperspectral image (HSI) captures a reflectance vector for every point on the ground. The corrected Indian Pines scene is composed of 200 spectral bands and spectrally similar classes that are also hard to separate at individual wavelengths [1]. The spectral data enable crop mapping, field inspection, and the targeting of areas for sampling, but pose statistical difficulties: labels are expensive; the number of pixels in each spectral class varies; and pixels within a field are correlated. Therefore, a practical crop classifier should be assessed on its ability to transfer to a set of images outside its training set, not on random pixels of a benchmark image. The development of hyperspectral classification has progressed from handcrafted spectral features to convolutional, recurrent, attention, and Transformer models [2], [3].

Early convolutional neural network (CNN) studies showed that feature learning achieves better performance than a shallow classifier; spectral-spatial feature learning provides even better performance [4]. Later approaches include pixel-pair learning [5], band-adaptive networks [6], 3-D networks with residuals [7], 3-D convolution [8], and hybrid 3-D/2-D convolutional networks [9]. Long-range spectral/spatial dependencies are captured by attention- and sequence-based models such as HSI-BERT [10], SpectralFormer [11], and spectral-spatial attention networks [12]. In situations where spatial context is dense and sufficient computational capacity exists, the above methods are effective; however, achieving high benchmark accuracy does not demonstrate transfer to an unseen agricultural field. The purpose of band selection is to retain original channels while minimizing acquisition, transmission, and interference costs. Techniques such as reconstruction-based ensembles with

Journal of Intelligent Engineering and Informatics, Volume 1, Issue 1, Pages 94-108
Corresponding author. E-mail address: lavanyas.bme@jerusalemengg.ac.in
Received 18 July 2026; **revised** 26 August 2026; **accepted** 31 August 2026; **published** 12 September 2026.

dual attention select informative bands [13], and measurement-learning techniques jointly learn a band mask together with a classifier [14]. Vegetation-informed selection [15] and superpixel-based selection [16] have been used for agricultural studies. However, the chosen small set may be biased due to the assignment of the evaluation split. A band can exhibit local field conditions if neighboring pixels in the same field are in both the training and test data sets. Random sampling of pixels, therefore, is susceptible to overestimating spectral-spatial performance because some training observations and test observations might be correlated.

To solve this problem, disjoint-region approaches have been used to partition spatial regions during evaluation [17]–[20]. A survey of 146 remote-sensing studies revealed that the use of correlation-aware sampling is limited and that partitioned or controlled sampling is more appropriate for assessing generalization [20]. A 75/25 stratified random split in the foundational workflow yields an optimistic score. This motivates a class-balanced logistic model for transfer between complete grass fields and the following research question: can two original spectral bands be identified that transfer between complete grass fields better than the fixed bands 66 and 76? Discriminating Grass-pasture is the main task in Indian Pines, while connected labeled regions are treated as field units. A single field per class is held out in each outer test, resulting in 12 field-pair evaluations. Band ranking, pair selection, standardization, and classifier fitting are performed using only the remaining training fields. There is a secondary 12-class experiment with 200 bands that is used to elucidate protocol effects from model-capacity effects. Given that all two-band candidates use the same linear learner, any improvement can only be due to band ranking and field-based validation, not hidden spatial context or other parameters. In FS-BPL, there are three interrelated changes. First, pixel-weighted separation is replaced by the stability score at a 'field' level, where a large field does not dominate the band ranking. Second, band pairs are chosen using nested complete-field validation that includes a penalty for the gap between the train and validation bands, band correlation, and band variability across folds.

Third and most importantly, the selected bands are applied in conjunction with the same class-balanced logistic classifier, thereby maintaining a linear decision boundary that is “transparent.” The contribution is not, therefore, any sort of new high-capacity classifier, but a field-aware selection and validation strategy to test transfer before increasing model complexity. In Section II, the scope and novelty of the research are defined. The mathematical model, architecture, and algorithm are given in Section III. The experimental approach and measurements are outlined in Section IV. All quantitative results, visual analyses, and ablations with failure cases are described in Sections V and VI. In Section VII, limitations of implementation and future validation are discussed. The novel central element is to consider, quantify, and show the gain, loss, and exposure of a two-band classifier when generalization is evaluated on whole fields rather than neighboring pixels.

II. Research Scope and Novelty

A. Study Scope and Boundaries

The calculations in this study were performed on the corrected Indian Pines cube (145×145×200) and its 16-class reference map [1]. The main task in the workflow remains the binary example of class 5 (Grass-pasture; 483 pixels) and class 6 (Grass-trees; 730 pixels). Logistic regression and class balancing are not affected. External scenes, image patches, transformed components, texture filters, or samples of any kind are not handed over to the proposed classifier. The spectral bands are numbered from 0 to 199 to correspond to the workflow.

The labeled map gives 8-neighbor connected components that make up field units. The two classes have 4 pasture components and 3 tree components, so there are $4 \times 3 = 12$ outer tests, each of which omits one component from each class. Evaluation groups are distinguished by ground-truth labels and are never used as predictor features. In each outer training set, complete units are again withheld by class for pair selection. The outer tests are weighted equally in the reported mean, and the standard deviation and range reflect field-to-field heterogeneity. Only classes containing at least 100 labeled pixels are included in the 12-class control and are used to compare random and spatially grouped evaluation.

B. Research Problem and Motivation

Several apparently good sections of the image may be obtained via random sampling of image pixels, in which neighboring training and test pixels have similar local field conditions and the local acquisition environment. The overall score thus may be related to the crop-specific spectral structure or to spatially correlated field conditions. It is important to note this distinction in Indian Pines, where the labeled crop areas are contiguous. For a deployed model, however, the model might be exposed to a different illumination regime, management regime, or parcel.

The experimental unit should thus be similar to the user's actual unit. This illustrates a second motivation: sensor cost. The 200-band system is flexible, but it is also fairly heavy in terms of acquisition data loads. If informative bands can be reliably transferred across fields, they may enable tight-scale measurements and low-resource edge inference. Selection of the bands thus need not be limited to fields scattered among fields containing different types of pixels. FS-BPL is designed to give the highest priority to field transfer while remaining small enough to be audited.

C. Methodological Novelty

The unique feature is the combination of field-aware band ranking, nested complete-field pair validation, and the minimum amount of necessary sensor data in the linear decision rule. For better class separation, FS-BPL includes a field mean regardless of the number of pixels. Three candidate pairs are then tested on the nested complete-field holdouts. The objective aims to encourage validation accuracy and discourage the train-validation gap, redundant bands, and unstable fold scores. Unlike other band-selection methods such as pixel-Fisher ranking [3], reconstruction-based band selection [13], and joint mask learning [14], FS-BPL embeds field transfer directly into the band-selection criterion.

This design can be separated into two aspects: interpretability and representational capacity. Selected wavelengths and the fitted logistic coefficients can be viewed, printed, and used without a spatial patch encoder. Finally, there are clear restrictions: FS-BPL does not model neighborhood texture, it cannot capture the representational power of full-spectrum models in every field, and FS-BPL must have representative training fields. However, differences in the classes, metrics, and partitions of the published studies prevent Table I from being used as a ranking of methods. In the above cases, for example, maps can be synthesized from 88.6% of the pixel samples with a uniformly distributed pattern in [16] or from 30 principal components to obtain an average accuracy of 88.42% in [19]. Using two bands in a binary test yields 68.4% balanced accuracy according to FS-BPL. Its superiority is demonstrated only against the matching fixed-pair reference under the same field-holdout method.

Table I. Protocol-aware comparison with recent Indian Pines studies.

Reference	Spectral input and classifier	Main partition	Reported Indian Pines result	Directly comparable?
[15]	45 selected bands + 2-D convolution	Pixel split; selected crop classes	95.97–99.35% class accuracy	No—different classes and spatial model
[14]	30 learned bands + 1-D convolution	Pixel sampling	89.08% overall accuracy	No—16 classes and different split
[19]	30 principal components + CNN/Transformer patches	Disjoint regions	88.42% average accuracy	No—16 classes and spatial patches
[16]	Crop-superpixel band subset + mapping classifier	Uniformly distributed pixels	88.6% mapping accuracy	No—16 classes and pixel sampling
FS-BPL	Two raw bands + balanced logistic regression	12 complete field-pair holdouts	68.4% balanced accuracy	Reference setting

III. Proposed Methodology

A. Mathematical Model

Let the hyperspectral cube be $X \in \mathbb{R}^{H\times W\times B}$, where H=145, W=145, and B=200. For each labeled target pixel n, the sample contains a spectrum $\mathbf{x}_n \in \mathbb{R}^B$, class $y_n \in \{5,6\}$, and coordinate $\mathbf{p}_n=(r_n,c_n)$. The binary dataset is defined by

$$D=\{(\mathbf{x}_n, y_n, \mathbf{p}_n)\}_{n=1}^{N}, \qquad N=1213. \tag{1}$$

where D denotes the binary dataset; x_n is the B-band spectrum of labeled pixel n; $y_n \in \{5, 6\}$ is its class label; $p_n = (r_n, c_n)$ is its row-column coordinate; n is the sample index; and N = 1213 is the total number of labeled samples in the binary task.

Using the dataset in (1), each class c is partitioned by an eight-neighbor connected-component operator into fields $G_c=\{g_{c,1},\ldots,g_{c,K_c}\}$. Two pixels share a field only when a path of horizontally, vertically, or diagonally adjacent pixels with class c connects them:

$$g(\mathbf{p}_n)=\mathrm{CC}_8(\mathbf{1}[y(\mathbf{p})=c]), \qquad K_5=4,\ K_6=3. \tag{2}$$

where $g(p_n)$ denotes the field/component assigned to the pixel at p_n; CC_8 is the eight-neighbor connected-component operator; $1[\cdot]$ is the indicator function forming the class-c mask; y_p is the class label at location p; c is the class under consideration; and K_5 = 4 and K_6 = 3 are the numbers of connected components for classes 5 and 6, respectively.

For the field partition in (2), the mean reflectance at band b for field g in class c is

$$m_{c,g,b}=\frac{1}{|g|}\sum_{n:\,\mathbf{p}_n} x_{n,b}\,. \tag{3}$$

where m(c,g,b) is the mean reflectance of field g in class c at band b; |g| is the number of labeled pixels in field g; the summation is over samples n whose coordinate p_n lies in g; and x_n,b is the reflectance of sample n at band b.

Using the field means in (3), FS-BPL ranks each band by the squared distance between class-level field means divided by the variability of field means within each class. A small constant ε prevents division by zero:

$$S_b=\frac{(\bar{m}_{5,b}-\bar{m}_{6,b})^2}{s_{5,b}^2+s_{6,b}^2+\varepsilon}, \qquad \bar{m}_{c,b}=\frac{1}{K_c}\sum_{G_c} m_{c,g,b}\,. \tag{4}$$

where S_b is the field-stability score for band b; m̄(c,b) is the mean of the field means for class c at band b; s(c,b) is the corresponding standard deviation of field means; ε is a small positive constant preventing division by zero; K_c is the number of fields in class c; G_c is the set of class-c fields; and m(c,g,b) is the field mean defined in (3), with c ∈ {5, 6}.

Bands are ranked using (4) from outer-training fields only. The 24 highest-scoring bands form the candidate set; bands 66 and 76 are retained as anchors so that the reference baseline cannot be excluded by ranking. For candidate pair (i,j), an inner fold holds out one complete training field from each class. Complete-field balanced accuracy (BA) in fold f is

$$\mathrm{BA}_f(i,j)=\frac{1}{2}\left(\mathrm{TPR}_{5,f}(i,j)+\mathrm{TPR}_{6,f}(i,j)\right). \tag{5}$$

where $\mathrm{BA}_{f(i,j)}$ is the balanced accuracy in inner fold f for candidate bands i and j; $\mathrm{TPR}_{5,f(i,j)}$ and $\mathrm{TPR}_{6,f(i,j)}$ are the class-wise true-positive rates (recalls) for classes 5 and 6 in fold f; f indexes the inner complete-field validation fold; and i and j denote candidate band indices.

The balanced-accuracy score in (5) is combined with the absolute train-validation gap G_f, the training-pixel pair correlation ρ_{ij}, and the standard deviation across inner folds σ_{BA}. The selected pair maximizes

$$Q(i,j)=\overline{\mathrm{BA}}(i,j)-0.20\,\overline{G}(i,j)-0.02\left|\rho_{ij}\right|-0.01\sigma_{\mathrm{BA}}(i,j). \tag{6}$$

where $Q(i,j)$ is the pair-selection objective; $\bar{\mathrm{A}}$B(i,j) is the mean balanced accuracy across inner folds; $\bar{G}(i,j)$ is the mean absolute training-validation gap across inner folds; $|\rho_{ij}|$ is the absolute training-pixel correlation between bands i and j; $\sigma_{\mathrm{BA}(i,j)}$ is the standard deviation of balanced accuracy across inner folds; and 0.20, 0.02, and 0.01 are the fixed penalty weights.

The objective in (6) keeps validation accuracy dominant while using small tie-breaking penalties for optimism, redundancy, and instability. After pair selection, each retained band is standardized from outer-training statistics:

$$z_{n,b}=\frac{x_{n,b}-\mu_b^{\mathrm{tr}}}{s_b^{\mathrm{tr}}}, \qquad b\in\{i^*,j^*\}. \tag{7}$$

where $z_{n,b}$ is the standardized value of sample n at band b; $x_{n,b}$ is its original reflectance; μ_b^{tr} and s_b^{tr} are the mean and standard deviation of band b computed from outer-training data only; b is one of the selected bands; and i^* and j^* are the selected optimal band indices.

The standardized inputs in (7) are passed to the class-balanced logistic model. The probability of Grass-trees is

$$P\left(y_n=6|\mathbf{z}_n\right)=\sigma\left(w_0+w_1 z_{n,i^*}+w_2 z_{n,j^*}\right), \qquad \sigma(a)=\frac{1}{1+e^{-a}}. \tag{8}$$

where $P(y_n = 6 \mid z_n)$ is the predicted probability that sample n belongs to class 6; y_n is the class label; z_n is the selected standardized feature vector; w_0 is the intercept and w_1 and w_2 are the fitted logistic coefficients; z_{n,i^*} and z_{n,j^*} are the standardized values of the selected bands; $\sigma(\cdot)$ is the logistic sigmoid; a is its scalar input; and e is Euler's number.

The probability model in (8) is fitted by minimizing weighted log loss with an L_2 term. The class weight $\alpha_c=N/(2N_c)$ makes the training contribution of both classes equal:

$$\mathrm{L}(\mathbf{w})=-\sum_{n\in\mathrm{D}_{\mathrm{tr}}}\alpha_{y_n}\left[y'_n\log p_n+(1-y'_n)\log\left(1-p_n\right)\right]+\frac{\lambda}{2}\|\mathbf{w}\|_2^2, \tag{9}$$

where L(w) is the weighted L2-regularized log-loss; w is the logistic coefficient vector; D_{tr} is the outer-training set; α_{y_n} is the class weight associated with sample n; $y'_n = 1[y_n = 6]$ is the binary target indicator; p_n denotes the probability $P(y_n = 6 \mid z_n)$ from (8) within this loss expression; λ is the L2 regularization coefficient; $\|w\|_2^2$ is the squared Euclidean norm of w; and log denotes the natural logarithm.

In (9), $y'_n=\mathbf{1}\left[y_n=6\right]$. Using the fitted parameters, the predicted class is obtained by the fixed probability threshold

$$\hat{y}_n=\begin{cases}6, & P\left(y_n=6|\mathbf{z}_n\right)\geq 0.5,\\ 5, & \text{otherwise.}\end{cases} \tag{10}$$

where $\hat{y}_n$ is the predicted class for sample n; $P(y_n = 6 \mid z_n)$ is the class-6 probability from (8); 0.5 is the fixed decision threshold; and labels 6 and 5 correspond to Grass-trees and Grass-pasture, respectively.

Using the prediction rule in (10), outer performance is summarized over every pasture-tree field pairing rather than over repeated pixels:

$$\overline{\mathrm{BA}}_{\text{field}}=\frac{1}{K_5K_6}\sum_{a=1}^{K_5}\sum_{b=1}^{K_6}\mathrm{BA}\left(g_{5,a},g_{6,b}\right). \tag{11}$$

where $\bar{\mathrm{BA}}_{\text{field}}$ is the mean balanced accuracy over all pasture-tree field pairs; K_5 and K_6 are the numbers of fields in classes 5 and 6; a and b index those fields; $g_{5,a}$ and $g_{6,b}$ denote the a-th class-5 field and b-th class-6 field; and $\mathrm{BA}[g_{5,a}, g_{6,b}]$ is the balanced accuracy for that held-out field pair.

B. System Architecture

Figure 1 summarizes the FS-BPL workflow. The pipeline proceeds from the HSI cube audit and field construction to stable band scoring, nested pair validation, linear crop classification, and reliability auditing. The lower branches identify the evaluation contract, pair-selection objective, interpretability check, and reporting stage. Only stage headings are retained in the diagram to keep the architecture concise.

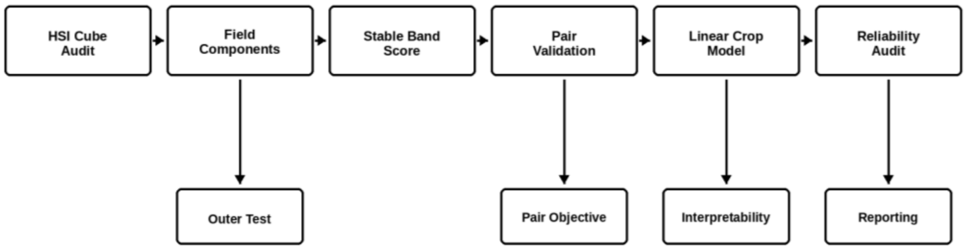

Figure 1. FS-BPL architecture.

C. Nested Validation Framework

Field construction is used only for evaluation and does not act as a spatial feature. Contiguous parcels in the reference map are identified from their class labels. In each outer fold, all pixels from one pasture component and one tree component are removed before any band score, scaler, or classifier is estimated. The remaining components are used for the inner complete-field validation. Thus, a labeled pixel cannot influence its own band selection through neighboring training pixels from the same field. This design is stricter than a random pixel split. The cross-product of components evaluates every pasture field against every tree field. Component sizes range from 18 to 318 pixels for pasture and from 222 to 270 pixels for trees. Balanced accuracy is therefore used as the primary metric, while ordinary accuracy is retained to reveal the effect of unequal test-field sizes.

D. Field-Stable Pair Selection

Pixel-Fisher ranking weights all pixels equally, allowing large fields to dominate class statistics. FS-BPL instead assigns each field one mean spectrum. The band score in (4) is high when class means are well separated and low when field-to-field variability is large. Candidate pairs are then evaluated with the objective in (6). The correlation penalty discourages redundant band pairs, while the train-validation gap penalizes pairs that fit the training fields well but transfer poorly. The fold-variability penalty prevents a pair from being selected because of one convenient validation component. The selected pair can therefore vary across outer folds as the available training fields change. This variation is intentional during algorithm validation, but a fixed two-channel sensor would require a consensus pair established across multiple scenes. Pair selection is therefore both a preprocessing step and an interpretability result.

E. Linear Classification and Decision Audit

No neighborhood patch is used by the classifier. The model is fitted on pixels from the outer training fields and applied unchanged to the held-out fields. The logistic boundary contains three fitted coefficients, so the direction and magnitude of each band's contribution remain visible. Row-normalized confusion matrices report class recall, while error maps preserve spatial location and show whether errors are concentrated near boundaries or extend across a shifted field. A 200-band logistic model provides an information-rich control, and depth-limited or unconstrained decision trees provide complexity ablations.

F. Algorithm 1

Table II summarizes the complete FS-BPL procedure. All selection operations remain inside the outer field split, and the algorithm refers directly to the numbered equations in Section III-A.

Table II. Algorithm 1—Field-Stable Band-Pair Learning (FS-BPL).

Step	Equation-referenced operation
Input	HSI cube, labels, coordinates, and anchor pair (66, 76).
1	Form field components using (2).
2	Compute field means using (3).
3	Rank bands using (4); retain the top 24 bands plus the anchor pair.
4	Evaluate candidate pairs on complete-field inner folds using (5).
5	Select the band pair by maximizing (6).
6	Standardize selected bands using (7) and fit the logistic model using (8) and (9).
7	Predict held-out field pixels using (10).
Output	Summarize field-level performance using (11); report accuracy, balanced accuracy, macro-F1, MCC, selected pairs, confusion matrices, and error maps.

IV. Experimental Design

A. Data and Class Protocol

The Indian Pines Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) scene contains 145 × 145 spatial locations and 200 bands after removal of water-absorption bands [1]. It contains 10,249 labeled pixels across 16 land-cover classes. Figure 2 shows the spectral composite and reference map. Because this is a measured scene rather than a synthetic cube, the spatial clustering of labels is preserved; this clustering motivates the use of complete-field rather than random-pixel evaluation.

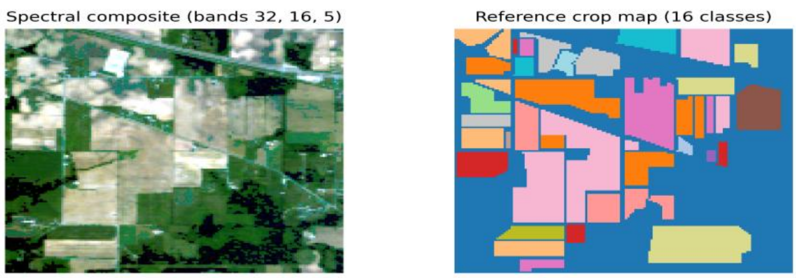

Figure 2. Indian Pines spectral composite and 16-class reference map.

Table III distinguishes the evaluation tasks and their corresponding protocols. The random binary reference uses a 75/25 stratified split with seed 4004 to reproduce the workflow baseline. The primary evaluation withholds one connected field from each target class, generating 12 outer tests. The secondary multiclass experiment uses the 12 classes with at least 100 labels and applies four grouped folds based on 12 × 12 spatial blocks; only out-of-fold predictions are evaluated.

Table III. Dataset and validation protocol used in each experiment.

Experiment	Classes	Features	Training/test rule	Purpose
Reference binary baseline	5 and 6	Bands 66, 76	Stratified random 75/25, seed 4004	Exact optimistic reference
Fixed-pair field test	5 and 6	Bands 66, 76	4 pasture × 3 tree component holdouts	Isolate split effect
FS-BPL	5 and 6	Two bands selected inside each outer fold	Same 12 component holdouts	Proposed evaluation
Full-spectrum control	5 and 6	200 bands	Same 12 component holdouts	Accuracy/interpret ability trade-off
Reference multiclass control	12 classes with ≥100 pixels	200 bands	Random 75/25	Reproduce stretch task
Spatial multiclass control	Same 12 classes	200 bands	Four grouped folds of 12 × 12 blocks	Out-of-fold spatial check

B. Model and Selection Settings

Standard scaling, class-balanced weights, and a fixed decision threshold are common characteristics of all logistic models. The full-spectrum control is identical except that the input model has 200 inputs. The 24 highest field-stability bands and the two anchors for the workflow are included in the candidate pool. These pair-objective weights are held constant (0.20, 0.02, and 0.01) and are not trained on the outer tests. For the decision-tree ablations, the seed is 4004, the maximum depth is set from 1 to 10, and there is an unconstrained tree. Primary results are based on all possible field holdouts rather than repeated random seeds.

C. Metrics and Statistical Reporting

The size of held-out fields can vary greatly, so accuracy can be skewed by a large sample. Therefore, the main measure is summarized across field pairs as a mean based on class recall; namely, the mean of the reciprocal of class recall is balanced accuracy. Macro-F1 scores are unweighted averages of the class-wise F1 scores. The Matthews correlation coefficient (MCC) is a measure of the correlation between predictions and class labels in the context of class imbalance. The unweighted mean and sample standard deviation (SD) are shown for the 12 outer field-pair tests, together with the full range in the ablation analysis. Confusion matrices are row-normalized. Because pixels within a field are not independent experimental units, pixel-level confidence intervals are not used to make significance claims.

D. Reproducibility Checks

The dimensions of the cube, the number of labels, and the number of target components are checked before training calculations, and these values are kept constant. Pixels outside the cube are never used in any outer training or test. Figures are created from stored or deterministic forecasts. Only out-of-fold predictions are shown for the 12-class spatial experiment. The pasture and tree components closest to one quarter of each class by size are selected by size and are not inspected for performance; therefore, they are not expected to form a visually appealing fold.

V. Results and Discussion

The mean spectra of Grass-pasture and Grass-trees are shown in Figure 3 along with the within-class variation. The class means vary in different parts of the spectrum, but there is considerable overlap in the variability bands. FS-BPL uses two spectral bands (visible range and near infrared), namely bands 19 and 76, in the size-matched outer fold. They are used in this fold, but not everywhere: only two of the 12 outer folds use the 19/76 pair. The stability obtained with the selected pairs themselves is an interesting result.

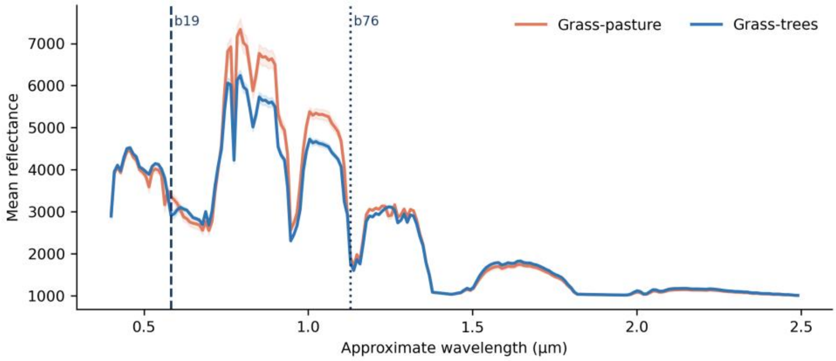

Figure 3. Mean target-class spectra with variability bands.

The main results of the binary classifications are shown in Table IV. Under "random pixel splitting," the fixed-pair baseline obtains a balanced accuracy of 0.791. When the same bands are evaluated on full fields, the mean BA drops to 0.582 ± 0.186 (-20.9 percentage points). This corresponds to the high impact of the evaluation split. A typical example is the field accuracy, with a mean of 0.757, where the larger tree parts of the field dominate the test pixel numbers; therefore, class-balanced reporting is needed.

Table IV. Main binary classification results.

Method	Test protocol	Features	Accuracy	Balanced accuracy	Macro-F1	MCC
Fixed-pair logistic	Random pixels	2	0.799	0.791	0.791	0.582
Fixed-pair logistic	Complete fields	2	0.757 ± 0.133	0.582 ± 0.186	0.571 ± 0.193	0.197 ± 0.387
FS-BPL	Complete fields	2	0.681 ± 0.221	0.684 ± 0.182	0.587 ± 0.226	0.282 ± 0.400
Full-spectrum logistic	Complete fields	200	0.812 ± 0.236	0.746 ± 0.201	0.705 ± 0.271	0.499 ± 0.419

FS-BPL shows a 10.2 percentage-point improvement in the mean balanced accuracy when compared with the fixed-pair field baseline. Its regular accuracy is lower because the selection objective does not focus on the most prevalent test pixels, but rather on both groups with equivalent weights. The increase does not apply only to recall or the macro-F1 score, as it is also seen in the improvements in MCC. The 200-band logistic control achieved a mean balanced accuracy (MBA) of 0.746, which is 6.2% higher than FS-BPL. Thus, as the number of bands in the spectral input to the sensor is reduced to two, there is a significant benefit for sensor simplicity and interpretability, but not all full-spectrum field-transfer benefit is maintained.

Figure 4 displays the holdout that was selected for the decision-plane audit based on size. Both the values of the fixed pair 66/76 and the selected pair 19/76 give a balanced accuracy of 0.495, but the pair 19/76 gives an accuracy of 0.973. This fold is only for visualization, and it was chosen for the size, not the performance, of the individual component fold. The held-out points are found in different regions of the spectral plane than the training points, which provides insight into the challenge of field shift as well as boundary fitting.

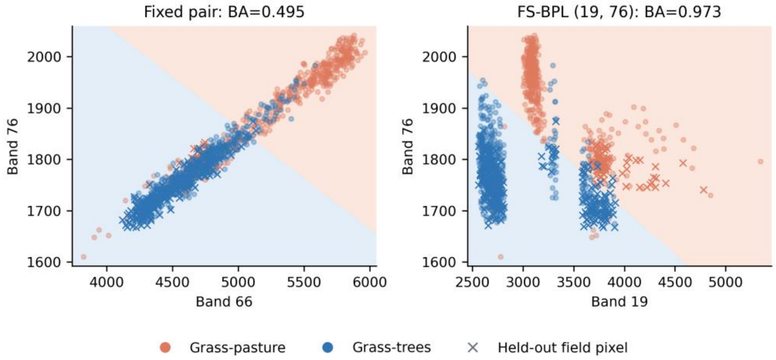

Figure 4. Fixed and FS-BPL decision planes.

The confusion matrices are compared in Figure 5 after normalization with respect to the number of rows. The relatively high recall of both classes is associated with random pixel mixing. The size-matched complete-field split is artificially inflated because it has a larger tree field and predicts almost all held-out pixels as Grass-trees (even though it only has an accuracy of ~0.5). FS-BPL restores recall for both classes. These predictions are shown on the map in Figure 6, where the confidence regions for the errors have the same coherence as those for the predictions themselves: the values are not taken at random from the field but occur within field components in the whole set.

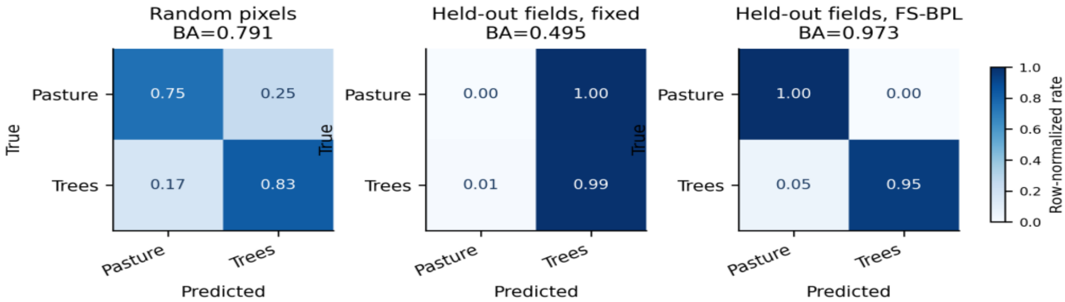

Figure 5. Row-normalized confusion matrices.

Figure 6. Spatial audit of the size-matched field partition.

Adding more features to the model does not eliminate the protocol bias. As illustrated in Figure 7, under random sampling, the accuracy of the trained models recovers to approximately one as the depth of the trees grows. However, under complete-field training, the accuracy of the trees degrades after shallow depths. The mean balanced accuracy of both models on the field-folded data is 0.650 at depth 3, whereas the unconstrained model achieves 0.626. Thus, the standardized linear rule (shallow tree) remains the most strongly defensible implementation, with deeper trees serving as diagnostic ablations.

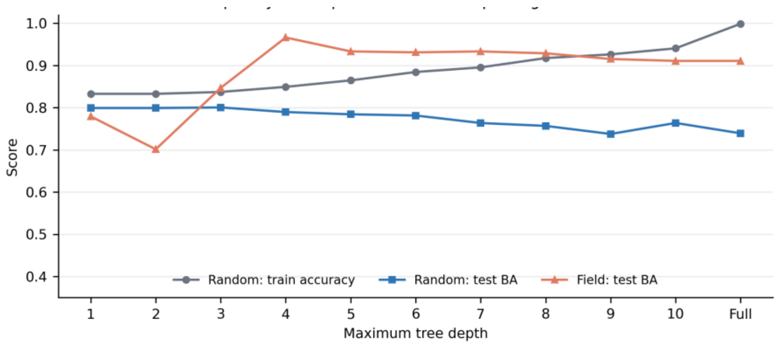

Figure 7. Decision-tree depth ablation under random and complete-field protocols.

The process of selecting bands is split into two stages, as Figure 8 shows. Field-stability scores are used to identify bands subject to the same level of variability in their field means while contributing to the same degree of separation among classes; the pair objective reorders candidates by taking into account both validation and stability across the fields. The difference between this and a learned latent mask is transparency, but there is uncertainty in the selection as well: multiple pairs may have similar objective values, and multiple selected pairs occur across the outer folds. It is therefore important not to consider a single pair to be a fixed sensor design without cross-scene validation and without taking several points into consideration.

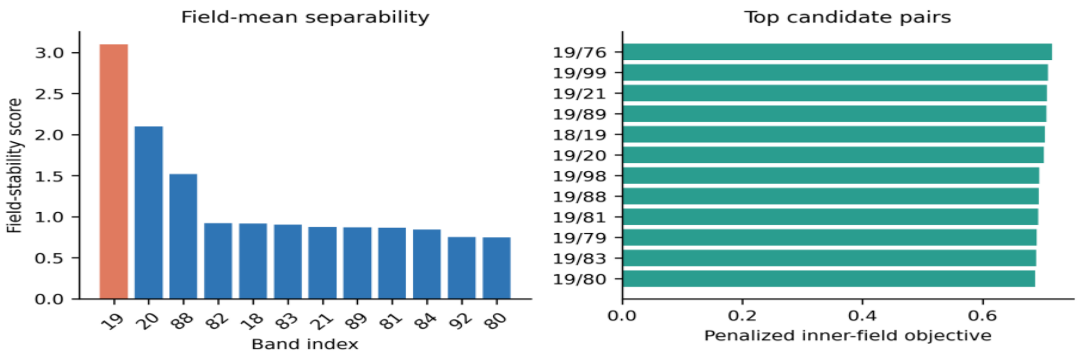

Figure 8. Field-stability band scores and top penalized pair objectives.

At a broader level of evaluation, the same protocol effect is observed in the 12-class control. As illustrated in Figure 9, the balanced accuracy is 0.878 under random pixel sampling and 0.805 under grouped (12 × 12) spatial-block evaluation, a drop of 7.3 percentage points under grouped (12 × 12) spatial-block evaluation. The reduction is in the same direction, but the decrease is smaller because neighboring blocks may still share both boundaries and environmental factors. Spectrally similar tillage classes of corn and soybean are more confounded. It can be seen in Figure 10 that the grouped errors correlate spatially, that is, they form clusters of errors, which is more important for practical purposes than isolated 'salt-and-pepper' errors because it might indicate that recalibration or field inspection is necessary.

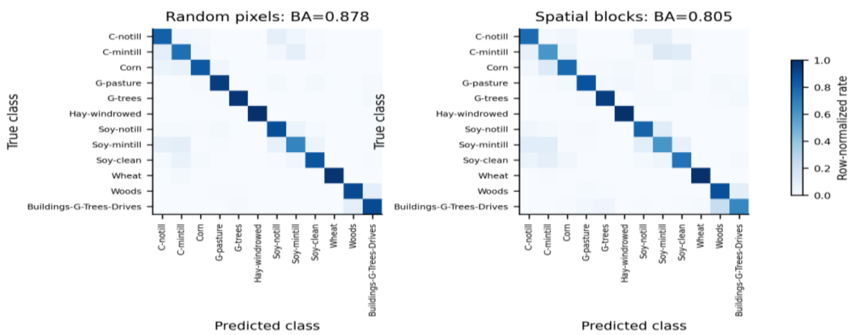

Figure 9. Twelve-class row-normalized confusion matrices.

Figure 10. Twelve-Class Reference and Spatial Prediction Maps.

Recent studies conducted on Indian Pines have reported higher percentages when using a larger number of bands, patches, or different class protocols [14]–[16], [19]. The values should not be directly compared with the 0.684 balanced accuracy reported here because the tasks, inputs, and partitions are diverse. FS-BPL shows greater gains than the fixed band-pair baseline under matching conditions, namely, the classifier, outer folds, and metrics. The supported claim is limited: field-aware pair selection benefits transfer over a matching fixed-pair reference, with the caveat that full-spectrum information and a broader surrounding dataset are still required for maximum transfer and deployment assurance.

VI. Ablation and Validation Analysis

The source of the gain is identified in Table V and Figure 11. Switching the fixed pair to a Pixel-Fisher pair results in a mean field-balanced accuracy of 0.513, showing that more powerful separation from the pixels does not necessarily yield greater field transfer. The score increases if the top two bands with the highest field stability are used: 0.647. Nesting adds complete-field pair validation, bringing it up to 0.684, showing that inner field validation, together with field-level ranking, adds to the overall effect.

Table V. Ablation results over the same 12 complete-field tests.

Variant	Features	Balanced accuracy (mean ± SD)	Change from fixed pair	Interpretation
Fixed bands 66/76	2	0.582 ± 0.186	—	Baseline reference
Pixel-Fisher pair	2	0.513 ± 0.152	−0.069	Pixel separation overfits fields
Field-score top two	2	0.647 ± 0.192	+0.065	Field means improve ranking
FS-BPL full selector	2	0.684 ± 0.182	+0.102	Inner field validation adds value
FS-BPL without class weights	2	0.689 ± 0.178	+0.107	Weighting is not the gain source
FS-BPL pair + depth-3 tree	2	0.650 ± 0.170	+0.068	Nonlinearity is less stable
FS-BPL pair + full tree	2	0.626 ± 0.149	+0.044	Extra complexity does not transfer
Full-spectrum logistic	200	0.746 ± 0.201	+0.164	More information, less sensor simplicity

This is not due to the class-weighting factor: without class weighting, a mean balanced accuracy of 0.689 is obtained, a value just larger than the 0.684 obtained by the balanced model. The reason for class weighting is not that it brings an improvement but that it assigns equal weight to both classes and fits the reference workflow. The use of nonlinear trees is not helpful for mean field transfer either. Different thresholds work for the intensity range of one field but do not work for another field when it changes. Thus, it seems that the most justifiable and explicable boundary is given by the two-band linear model. Both mean ± one SD and individual field-pair scores are given in Figure 11. The FS-BPL range is between 0.340 and 0.973. These results show that the three tests with the highest percentage of pasture showed better performance with the fixed pair than with FS-BPL, where pasture was not available for differentiation during training. Although it decreases the average leakage bias, FS-BPL does not prevent the occurrence of agricultural domain shift.

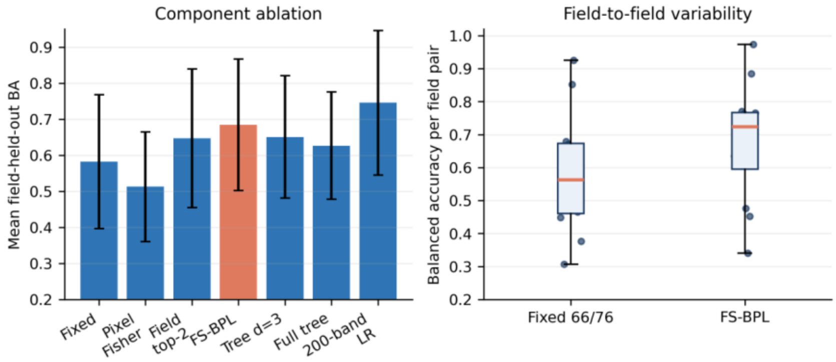

Figure 11. Component ablation and field-to-field score distributions.

Four concrete properties are used to assess interpretability: two measurements of the spectrum—the number of lines and the proportion of the lines present—plus preservation of the identity of the wavelength that can be used to specify a particular line within the spectrum and the linear decision boundary that would be operated in the case of a spatially referenced error. Component contributions are evaluated by fixed-pair, Pixel-Fisher, field_score, pair-objective, class-weight, and tree-depth types. Reproducibility is facilitated by deterministic outer folds, a separate random baseline, out-of-fold multiclass maps, and saved fold-level metrics. Together, these checks enable component performance to be assessed and explain what is not working in the method.

VII. Implementation and Future Directions

FS-BPL can be used in low-resource inference. Each pixel requires only two calibrations, two simple sums, and a threshold; the straight separating hyperplane can be tested without explicitly computing the sigmoid. This is much lighter than inference with patch-based convolution or Transformer models [9]–[12] or [19]. If a full hyperspectral instrument is already available, the 200-band logistic control is also feasible and yields higher mean balanced accuracy. The two-band difference is greatest when acquisition bandwidth, transmission, calibration, and interpretability costs dominate. Four limitations remain. First, the historical scene cannot establish transfer, seasonality, or geographic generalization. Second, connected label parts do not exactly correspond to actual field parcels; it would be better to have parcel polygons. Third, the selected pair is not necessarily the same for each fold, which makes a permanent two-filter instrument unavailable until sufficient evidence is available.

Fourth, only a few fields are available for inner validation, giving a high selection variance, manifested by poor validation of the largest pasture component. This calls for future validation with multiple scenes/dates and multiple farms, definition of a target field before model evaluation, and a possibly optimized consensus pair value across farms with a small calibration difference. Discard fields not at the extremes of the training distribution; do not fit them into a class. Distributionally robust objectives may focus on worst-field loss, not mean field loss, and sensor-aware constraints may include minimum spectral spacing and realistic filter bandwidth. The pair and coefficients should then be 'frozen' before a potential campaign. A prospective split for another deployment-readiness evaluation should be considered, rather than another split based on a random factor.

VIII. Conclusion

The issue of spatial dispersion is addressed by FS-BPL through the definition of the concern in terms of field-level selection and evaluation. The method adopts two raw bands that are selected on the basis of their stability in field means, with complete-field validation nested within the procedure, while retaining the standardized class-balanced logistic classifier. Aiming for balanced accuracy, FS-BPL generates a 10.2 percent accuracy gain across 12 pasture-tree field holdouts using only two of 200 channels. The full-spectrum logistic control remains 6.2 percentage points higher, quantifying the cost of sensor simplicity. The translatability of ablations with deeper decision trees and Pixel-Fisher ranking is lower. The outcome is a pure band-selection objective, a full validation protocol for the entire field, and an explicit failure audit. Future studies should define a consensus band pair before application on the farm and validate it throughout the entire season and across different types of sensors.

References

[1] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, “220 band AVIRIS hyperspectral image data set: June 12, 1992 Indian Pine Test Site 3,” Purdue University Research Repository, 2015. doi: 10.4231/R7RX991C.

[2] N. Audebert, B. Le Saux, and S. Lefèvre, “Deep learning for classification of hyperspectral data: A comparative review,” IEEE Geosci. Remote Sens. Mag., vol. 7, no. 2, pp. 159–173, Jun. 2019. doi: 10.1109/MGRS.2019.2912563.

[3] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, “Deep learning classifiers for hyperspectral imaging: A review,” ISPRS J. Photogramm. Remote Sens., vol. 158, pp. 279–317, Dec. 2019. doi: 10.1016/j.isprsjprs.2019.09.006.

[4] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 10, pp. 6232–6251, Oct. 2016. doi: 10.1109/TGRS.2016.2584107.

[5] W. Li, G. Wu, F. Zhang, and Q. Du, “Hyperspectral image classification using deep pixel-pair features,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 2, pp. 844–853, Feb. 2017. doi: 10.1109/TGRS.2016.2616355.

[6] A. Santara, K. Mani, P. Hatwar, A. Singh, A. Garg, K. Padia, and P. Mitra, “BASS Net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 9, pp. 5293–5301, Sep. 2017. doi: 10.1109/TGRS.2017.2705073.

[7] Z. Zhong, J. Li, Z. Luo, and M. Chapman, “Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 2, pp. 847–858, Feb. 2018. doi: 10.1109/TGRS.2017.2755542.

[8] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben Amar, “3-D deep learning approach for remote sensing image classification,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 8, pp. 4420–4434, Aug. 2018. doi: 10.1109/TGRS.2018.2818945.

[9] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, “HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification,” IEEE Geosci. Remote Sens. Lett., vol. 17, no. 2, pp. 277–281, Feb. 2020. doi: 10.1109/LGRS.2019.2918719.

[10] J. He, L. Zhao, H. Yang, M. Zhang, and W. Li, “HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from transformers,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 1, pp. 165–178, Jan. 2020. doi: 10.1109/TGRS.2019.2934760.

[11] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, “SpectralFormer: Rethinking hyperspectral image classification with transformers,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–15, 2022, Art. no. 5518615. doi: 10.1109/TGRS.2021.3130716.

[12] H. Sun, X. Zheng, X. Lu, and S. Wu, “Spectral–spatial attention network for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 5, pp. 3232–3245, May 2020. doi: 10.1109/TGRS.2019.2951160.

[13] S. K. Roy, S. Das, T. Song, and B. Chanda, "DARecNet-BS: Unsupervised dual-attention reconstruction network for hyperspectral band selection," IEEE Geosci. Remote Sens. Lett., vol. 18, no. 12, pp. 2152–2156, Dec. 2021. doi: 10.1109/LGRS.2020.3013235.
[14] C. O. Ayna, R. Mdrafi, Q. Du, and A. C. Gurbuz, "Learning-based optimization of hyperspectral band selection for classification," Remote Sens., vol. 15, no. 18, Art. no. 4460, Sep. 2023. doi: 10.3390/rs15184460.
[15] L. Agilandeeswari, M. Prabukumar, V. Radhesyam, K. L. N. B. Phaneendra Kumar, and A. Farhan, "Crop classification for agricultural applications in hyperspectral remote sensing images," Appl. Sci., vol. 12, no. 3, Art. no. 1670, Feb. 2022. doi: 10.3390/app12031670.
[16] Y. Tan, J. Gu, L. Lu, L. Zhang, J. Huang, L. Pan, Y. Lv, Y. Wang, and Y. Chen, "Hyperspectral band selection for crop identification and mapping of agriculture," Remote Sens., vol. 17, no. 4, Art. no. 663, Feb. 2025. doi: 10.3390/rs17040663.
[17] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 55, no. 2, pp. 862–880, Feb. 2017. doi: 10.1109/TGRS.2016.2616489.
[18] X. Cao, C. Li, J. Feng, and L. Jiao, "Semi-supervised feature learning for disjoint hyperspectral imagery classification," Neurocomputing, vol. 526, pp. 9–18, Mar. 2023. doi: 10.1016/j.neucom.2023.01.054.
[19] P. Zhang, H. Yu, P. Li, and R. Wang, "TransHSI: A hybrid CNN-Transformer method for disjoint sample-based hyperspectral image classification," Remote Sens., vol. 15, no. 22, Art. no. 5331, Nov. 2023. doi: 10.3390/rs15225331.
[20] K. T. Decker and B. J. Borghetti, "A survey of sampling methods for hyperspectral remote sensing: Addressing bias induced by random sampling," Remote Sens., vol. 17, no. 8, Art. no. 1373, Apr. 2025. doi: 10.3390/rs17081373.