Hyperspectral Band Selection for Ground Fuel Classification for Prescribed Fires

Mahmad Isaq Karankot , Ethan M.Glenn , Muhammad Umer Masood , Xiaobing Zhou , Bradley M. Whitaker *

Posted Date: 3 March 2026

doi: 10.20944/preprints202603.0211.v1

Keywords: hyperspectral imaging; band selection; machine learning; prescribed fire

 doi:10.20944/preprints202603.0211.v1

Article

Hyperspectral Band Selection for Ground Fuel Classification for Prescribed Fires

Mahmad Isaq Karankot [1,*], Ethan M.Glenn [1], Muhammad Umer Masood [2], Xiaobing Zhou [2] and Bradley M. Whitaker [1]

[1] Electrical and Computer Engineering, Montana State University, Bozeman, MT 59717, USA
[2] Geological Engineering Department, Montana Technological University, Butte, MT, USA
* Correspondence: mahmadisaq.karankot@student.montana.edu

Abstract

Hyperspectral image (HSI) analysis plays a central role in remote sensing tasks requiring fine-grained material discrimination, vegetation health assessment, and post-disturbance monitoring. Yet, the high dimensionality and strong spectral redundancy in HSIs often reduce the efficiency and reliability of machine learning models. These challenges are especially important in wildfire science and prescribed-fire monitoring, where spectral responses vary due to burn severity, char deposition, canopy structure, and early vegetation recovery. Benchmark datasets such as Indian Pines and Pavia University provide controlled environments for algorithm evaluation, but real-world post-fire forest conditions pose additional complexity. This study presents a unified and comprehensive evaluation of four band-selection strategies: Principal Component Analysis (PCA), Spatial–Spectral Edge Preservation (SSEP), Spectral-Redundancy Penalized Attention (SRPA), and a Deep Reinforcement Learning (DRL)–based selector. These strategies are combined with classical machine learning and deep learning classifiers: Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and 3D Convolutional Neural Networks (3D-CNN). The full pipeline includes exploratory data analysis, preprocessing, patch-based spatial–spectral modeling, consistent train–validation protocols, and multi-dataset evaluation across Indian Pines, Pavia University, and a new custom VNIR hyperspectral dataset collected after prescribed burns at the Lubrecht Experimental Forest in Montana, USA. By systematically comparing statistical, edge-aware, attention-guided, and reinforcement-learning-based band-selection strategies, this work identifies compact yet informative spectral subsets that enhance classification performance while reducing computational cost. Importantly, the inclusion of the Montana prescribed-burn dataset provides a unique real-world testbed for understanding band-selection behavior in fire-affected forest environments. Overall, this study contributes a generalizable and extensible framework for HSI dimensionality reduction and classification, laying the groundwork for future applications in wildfire assessment, vegetation recovery monitoring, and remote sensing.

Keywords: hyperspectral imaging; band selection; machine learning; prescribed fire

1. Introduction

Hyperspectral imaging (HSI) has emerged as a powerful remote sensing tool, enabling fine-grained discrimination of materials, vegetation health assessment, and monitoring of environmental disturbances due to its hundreds of contiguous narrow spectral bands [1–3]. The rich spectral information supports applications ranging from land cover classification [4] and crop mapping [5] to soil analysis [6] and quality inspection in agriculture [7]. However, the high dimensionality of HSI data introduces significant challenges, including the "curse of dimensionality," spectral redundancy, computational complexity, and reduced model generalization in machine learning pipelines [8,9].

Band selection techniques address these issues by identifying a compact, informative subset of bands that preserve discriminative power while minimizing redundancy and noise [1,10]. Traditional

 doi:10.20944/preprints202603.0211.v1

approaches, such as Principal Component Analysis (PCA), rely on variance maximization but often fail to capture task-specific discriminability [3]. More advanced methods incorporate spatial-spectral information [11], attention mechanisms [12], or graph-based structures [13]. Recently, deep reinforcement learning (DRL) has shown promise for adaptive, task-driven band selection in HSI classification, outperforming conventional techniques in semisupervised and unsupervised scenarios [14–17].

These challenges are particularly acute in wildfire science and prescribed fire management, where spectral responses are influenced by burn severity, char deposition, canopy structure, and early vegetation recovery [18,19]. Wildfires pose a growing threat to ecosystems, communities, and economies in the western United States, with Montana experiencing particularly pronounced impacts due to its vast forested landscapes, dry continental climate, and increasing human development in the wildland-urban interface. Over the past two decades, Montana has seen a dramatic rise in wildfire activity: average annual acres burned have exceeded 300,000 since 2000–more than ten times the pre-1950 average of less than 30,000 acres–driven by warmer temperatures, prolonged droughts, fuel accumulation from historical fire suppression, and climate change-induced aridity [20,21]. Recent seasons illustrate variability but underscore the trend: in 2024, over 352,000 acres burned, while 2025 saw a milder year with approximately 76,000 acres (one of the fourth-lowest totals in the last 15 years), attributed to favorable weather but still highlighting ongoing risks [22,23].

Prescribed fires–carefully planned, low-intensity burns–serve as a critical proactive tool to mitigate these threats by reducing hazardous fuel loads, restoring ecological processes, improving wildlife habitat, recycling nutrients, and promoting resilient forests. In Montana, where historical natural fire regimes burned tens to hundreds of thousands of acres annually before suppression, prescribed burning helps counteract fuel buildup and supports forest health. State and federal programs (e.g., DNRC, USFS Helena-Lewis and Clark National Forest) have expanded efforts, with recent accomplishments including tens of thousands of acres treated annually and proposals to scale to approximately 40,000 acres per year through 2045 on national forests alone [24–26]. Despite these advances, challenges persist, including limited burn windows due to air quality regulations, smoke concerns, and the need for precise monitoring of fuel types and post-burn recovery.

Accurate ground fuel classification–distinguishing live vegetation, dead biomass, and species-specific conditions–is essential for planning effective prescribed burns, predicting fire behavior, and evaluating ecological outcomes. Hyperspectral remote sensing, with its fine spectral resolution, offers significant potential for this task, yet real-world applications in fire-affected Montana forests remain underexplored due to data scarcity and dimensionality issues [27–29].

Despite advances in HSI band selection for specific tasks [30,31], few studies integrate multiple strategies (e.g., statistical, edge-aware, attention-guided, and reinforcement learning-based) in a unified pipeline, especially for prescribed fire applications. Prior work has explored attention and edge-aware methods for burned vegetation classification [32] and self-supervised/DRL approaches for prescribed burn impact analysis [33], highlighting the potential of learning-based techniques in fire-affected scenes. However, systematic comparisons across benchmarks and real-world Unmanned Aircraft System (UAS) datasets remain limited, particularly for ground fuel mapping in controlled burns.

This study addresses these gaps by presenting a comprehensive evaluation of four band-selection strategies–Principal Component Analysis (PCA), Spatial-Spectral Edge Preservation (SSEP), Spectral-Redundancy Penalized Attention (SRPA), and Deep Reinforcement Learning (DRL)–combined with classical (Random Forest, Support Vector Machines, K-Nearest Neighbors) and deep (3D Convolutional Neural Networks) classifiers. We apply a modular, reproducible pipeline to benchmark datasets (Indian Pines, Pavia University, Salinas, Botswana, Kennedy Space Center) and a novel VNIR hyperspectral dataset collected via UAS over prescribed burn sites at the Lubrecht Experimental Forest, Montana, USA. This dataset captures pre-burn conditions in a controlled thin-burn plot, providing a unique real-world testbed for understanding band-selection behavior in fire-prone forest environments.

By emphasizing exploratory data analysis (EDA), preprocessing, patch-based modeling, and cross-dataset validation, our work identifies compact spectral subsets that enhance classification

 doi:10.20944/preprints202603.0211.v1

accuracy while reducing computational demands. The inclusion of the Montana dataset bridges the gap between controlled benchmarks and operational remote sensing in wildfire contexts, contributing a generalizable framework for HSI dimensionality reduction and classification with applications in prescribed fire monitoring, vegetation recovery assessment, and fuel load management.

2. Materials and Methods

The goal of this study is to systematically evaluate a unified and reproducible hyperspectral image (HSI) classification pipeline across both benchmark airborne datasets and a complex UAS-based visible-near-infrared (VNIR) dataset captured over prescribed burn sites in the Lubrecht Experimental Forest, a large outdoor forest research laboratory located in the Blackfoot River drainage, about 30 miles northeast of Missoula, Montana. To accomplish this, we designed a modular workflow consisting of four tightly integrated stages illustrated in process diagram Figure 1: (1) dataset preparation and exploratory data analysis (EDA), (2) noise detection and data cleaning, (3) band-selection using classical and deep-learning-based feature reduction techniques, and (4) classification using both traditional machine learning models and deep learning architectures. All components were implemented in a consistent, dataset-agnostic fashion so that the same pipeline could be applied to Indian Pines, Pavia University, Salinas, Botswana, KSC, and the Montana VNIR dataset without modification.

This section describes the materials, datasets, preprocessing steps, algorithms, model architectures, and evaluation design used in the study, presenting each part as a connected narrative that reflects the logical progression of the project.

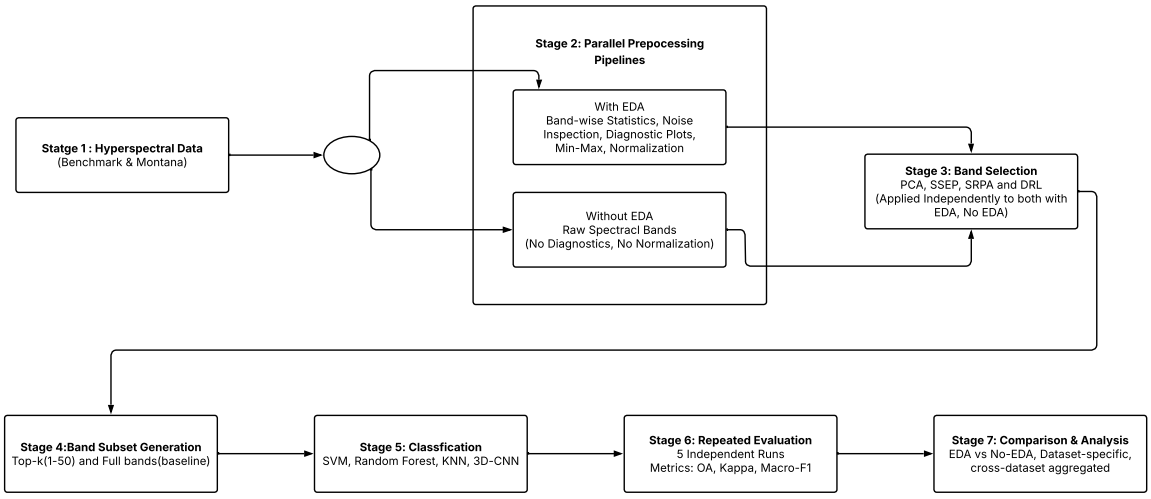

Figure 1. Unified workflow for hyperspectral image classification, illustrating data preparation, exploratory data analysis, band selection, and classification stages applied consistently across all datasets

2.1. Data

Hyperspectral imagery consists of hundreds of narrow and contiguous spectral bands, providing rich spectral information for material discrimination. However, this high spectral resolution also introduces significant redundancy and the well-known Hughes phenomenon (also known as the "curse of dimensionality"), motivating the need for dimensionality reduction and band selection prior to classification. Consequently, standardized preprocessing and band selection are widely adopted in hyperspectral image analysis pipelines [1–3].

We evaluate a unified hyperspectral classification framework using a diverse collection of datasets that span controlled benchmark scenes and a complex real-world unmanned aircraft system (UAS)–based acquisition. Specifically, experiments were conducted on five widely used benchmark hyperspectral datasets—Indian Pines, Pavia University, Salinas Valley, Botswana, and Kennedy Space Center [34]—as well as a custom visible–near-infrared (VNIR) hyperspectral dataset collected over prescribed burn sites in Montana, USA. The benchmark datasets represent a range of land-cover

 doi:10.20944/preprints202603.0211.v1

types, spatial resolutions, sensors, and class distributions, and have been extensively used in prior hyperspectral classification and band-selection studies, enabling meaningful comparison with existing literature [8,9,35].

A summary of the benchmark datasets, including sensor type, spatial dimensions, number of spectral bands before and after sensor-documented noisy-band removal, number of classes, and spatial resolution, is provided in Table 1. The cleaned benchmark cubes served as inputs to the subsequent exploratory data analysis. Rather than treating each benchmark dataset independently, all were processed using the same unified and dataset-agnostic pipeline to ensure consistency across preprocessing, band selection, and classification stages. This design allows performance differences to be attributed to algorithmic choices rather than dataset-specific handling.

In contrast to the benchmark scenes, the Montana VNIR dataset represents a real-world post-fire forest environment acquired using a UAS platform, characterized by higher spectral noise, severe class imbalance, sparse ground truth, and complex canopy structure. Detailed acquisition parameters, preprocessing steps, and dataset-specific challenges for the Montana dataset are described separately in Section 2.1.1, as they differ substantially from those of the benchmark datasets and require specialized treatment.

Table 1. Summary of hyperspectral datasets used in this study.

Dataset	Sensor	Spatial Dim. (H × W)	Bands (Orig./Cleaned)	Classes	Spatial Res. (m)
Indian Pines	AVIRIS	145 × 145	224 / 200	16	20
Pavia University	ROSIS	610 × 340	103 / 103	9	1.3
Salinas	AVIRIS	512 × 217	224 / 204	16	3.7
Botswana	Hyperion	1476 × 256	242 / 145	14	30
KSC	AVIRIS	512 × 614	224 / 176	13	18

2.1.1. Montana

The hyperspectral images used in this study were acquired on 11 May 2024 over a controlled thin-burn plot (330 m × 300 m) in the Lubrecht Experimental Forest, approximately 42 km east of Missoula, Montana, USA. Data collection was conducted using a Headwall co-aligned visible–near infrared (VNIR) and short-wave infrared (SWIR) imaging system, integrated with a FreeFly ALTA X heavy-duty unmanned aircraft system (UAS). Prior to the flight, the imaging system was calibrated using a certified Spectralon white diffuse reflectance standard to determine the optimal exposure time, frame period, and recommended flight speed. Based on this information, a flight plan was designed to maintain a nearly constant altitude relative to the terrain, thereby eliminating the need for terrain correction in the resulting imagery. Magnetometer calibration was also performed to ensure accurate detection of the Earth's magnetic field for reliable navigation. The precise location of the imaging system was recorded using a Trimble SPS585 global navigation satellite system (GNSS) unit for Real-Time Kinematic (RTK) surveying. Before executing the flight plan, the Inertial Measurement Unit (IMU) was initialized to track the aircraft's movement and orientation. The hyperspectral imaging system captured 543 spectral band images — 273 in the VNIR range (400–1000 nm) and 270 in the SWIR range (900–2500 nm). The data files acquired were preprocessed using Spectral View, the proprietary software provided by the manufacturer of the hyperspectral imaging system. The raw data, initially recorded in digital numbers (DNs), was corrected for noise and dark current using imagery collected prior to data acquisition. This process converted the DN values into spectral radiance (mW/(cm²·sr·µm)). The radiance data was then transformed into spectral reflectance using a calibrated reference tarp placed within the imaging scene during processing. Finally, the processed images were exported in GeoTIFF format to enable further analysis using standard image processing software.

 doi:10.20944/preprints202603.0211.v1

2.2. Exploratory Data Analysis

Exploratory data analysis (EDA) was conducted prior to model training to examine radiometric distributions, assess spectral redundancy, and identify noisy or unstable bands. EDA plays a critical role in ensuring data integrity and preventing the propagation of artifacts into downstream band selection and classification model [9,27].

2.3. EDA for Benchmark Hyperspectral Datasets

The objective of the benchmark dataset EDA was to standardize all preprocessing steps across the five classical hyperspectral scenes—Indian Pines, Pavia University, Salinas, Botswana, and Kennedy Space Center (KSC)—so that band–selection algorithms could be evaluated under consistent noise conditions. These scenes differ substantially in spatial dimensions, radiometric behavior, and class distributions, making a unified and statistically grounded EDA pipeline necessary. The goal of this pipeline was to verify data integrity, characterize radiometric distributions, identify low–SNR or water–absorption bands, and construct noise–free normalized cubes to serve as inputs for SSEP, SRPA, and DRL–based band selection as well as for classical and deep classifiers.

Each dataset was loaded from its `.mat` hyperspectral file and associated ground–truth (gt) `.mat` label mask using dataset–specific variable keys. Upon loading, the benchmark hyperspectral cubes exhibited the following dimensions: Indian Pines ($145 \times 145 \times 200$), PaviaU ($610 \times 340 \times 103$), Salinas ($512 \times 217 \times 204$), Botswana ($1476 \times 256 \times 145$) and KSC ($512 \times 614 \times 176$). Basic integrity diagnostics included computation of global intensity minima and maxima, band–wise means, and band-wise standard deviations. Botswana displayed reflectance values ranging from 0 to 45106, with an average band-wise mean of 1749.62 and an average standard deviation of 420.48. Indian Pines ranged from 955 to 9604, with an average band–wise mean 2652.39 and an average standard deviation 336.36. KSC ranged from –27 to 1244, with average band–wise mean 93.42 and an average standard deviation of 72.36. PaviaU showed intensities between 0 and 8000, with an average band-wise mean of 1389.12 and an average standard deviation of 716.43. Salinas ranged from –11 to 9207, with corresponding mean band-wise mean of 1196.40 and an average standard deviation of 389.91. These statistics confirmed proper radiometric scaling and the absence of corruption in the raw reflectance values.

Ground-truth label distributions were analyzed to verify class presence and quantify class imbalance. Botswana contained 15 classes with pixel counts ranging from 95 to 314, against 374,608 background pixels. Indian Pines exhibited 16 classes, including extremely sparse categories (e.g., class 1 with only 46 pixels) and dense categories such as class 11 with 2455 pixels. KSC contained 13 classes with class populations ranging from 105 to 927 and background comprising 309,157 pixels. PaviaU included 9 semantic classes, with some highly populated categories (e.g., class 2 with 18,649 pixels) and background containing 164,624 pixels. Salinas exhibited 16 classes, with major categories (e.g., class 8 with 11,271 pixels) and small categories (e.g., class 13 with 916 pixels). These results confirm substantial class imbalance across all benchmark datasets, which should be considered in downstream learning.

To characterize spectral noise, each hyperspectral cube was reshaped into a matrix of size $N \times B$, where $N = H \cdot W$, enabling band–wise analysis across all spatial pixels. For each spectral channel b, we computed its minimum and maximum intensities (m_b, M_b), mean μ_b, standard deviation σ_b, dynamic range $R_b = M_b - m_b$, and zero–fraction

$$f_b^{(0)} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}[X_{i,b} = 0] \tag{1}$$

Bands were flagged as suspicious if they exhibited extremely low variance (i.e., σ_b lying in the lowest 5% percentile of all variances), unusually small dynamic range (i.e., R_b less than 10^{-3} of the global range), or high zero–fraction (i.e., $f_b^{(0)} \geq 0.80$). Although none of the benchmark datasets contained high zero–fraction bands, each dataset presented a small set of low–variance bands that aligned with known water–absorption and sensor–instability regions. This EDA-based

 doi:10.20944/preprints202603.0211.v1

noisy-band identification was performed in addition to the initial sensor-documented band exclusion summarized in Table 1. The heuristic procedure identified the following candidate noisy bands: Botswana—[111, 138–144]; Indian Pines—[102–104, 143–145, 195, 197–199]; KSC—[0–8]; PaviaU—[0–5]; Salinas—[106–109, 146–148, 198, 201–203]. These empirically identified indices match previously documented noisy–band lists in the hyperspectral imaging literature.

Based on these technical diagnostics and literature support, we removed the identified noisy bands from each dataset, yielding cleaned spectral dimensionalities of 137 bands (Botswana), 190 bands (Indian Pines), 167 bands (KSC), 97 bands (PaviaU), and 193 bands (Salinas). Spatial dimensions were unchanged, ensuring that only the spectral axis was compressed. This removal eliminated known water–absorption wavelengths and low–SNR regions while preserving all meaningful spectral–spatial structure required for downstream tasks.

After noisy–band removal, each dataset was normalized per spectral channel using min–max scaling:

$$X_{i,b}^{\text{norm}} = \frac{X_{i,b} - m_b}{M_b - m_b + \varepsilon} \tag{2}$$

where ε prevents division by zero. This transformation places all spectral channels in the $[0, 1]$ range, eliminates scale disparities among bands, and stabilizes distance metrics and gradients for machine learning models. The cleaned and normalized hyperspectral cubes and associated summary statistics were retained for downstream analysis.

The results of this benchmark EDA demonstrate that the unified pipeline not only confirms the structural and radiometric consistency of all datasets but also quantitatively identifies and removes noisy spectral regions using both heuristic and literature–validated criteria. By producing noise–controlled hyperspectral cubes with well–characterized statistics, the benchmark EDA ensures that subsequent band–selection and classification experiments are conducted under standardized, reproducible, and scientifically rigorous conditions.

2.4. EDA for the Montana UAV Dataset

The Montana VNIR dataset consists of a drone-acquired hyperspectral mosaic with 273 spectral bands and spatial dimensions of 10706 × 8360 pixels. Before applying any band-selection algorithms or training classification models, we conducted a comprehensive exploratory data analysis (EDA) to (i) construct an accurate pixel-wise ground truth (GT), (ii) assess class distribution and label sparsity, and (iii) evaluate spectral-band radiometric quality and class-wise separability.

2.4.1. Ground-Truth Construction and Alignment Verification

Field plot measurements, provided as longitude–latitude points with species labels, were converted into a pixel-wise raster mask using a custom reprojection and rasterization pipeline. The GPS coordinates (EPSG:4326) were transformed into the coordinate reference system of the VNIR cube using `rasterio.warp.transform`, followed by conversion to pixel indices via the GeoTIFF affine transform. Each species was assigned a unique integer class ID (with 0 reserved for background), and points within the VNIR footprint were written into a 2-D mask saved as `montana_gt.npy`.

To validate GT–image alignment, we generated a pseudo-RGB composite generated using three high-SNR VNIR bands (R=150, G=100, B=50) and overlaid with the GT to verify geometric alignment. This procedure revealed and resolved an initial misalignment between labels and imagery. After reconstructing the GT using accurate reprojection, the overlay confirmed correct spatial correspondence between labeled pixels and canopy/burned regions. As shown in Figure 2, the corrected GT aligns well with canopy structures visible in the hyperspectral mosaic. This verification step is essential, as even small registration errors in UAS hyperspectral imagery can significantly degrade classification accuracy.

 doi:10.20944/preprints202603.0211.v1

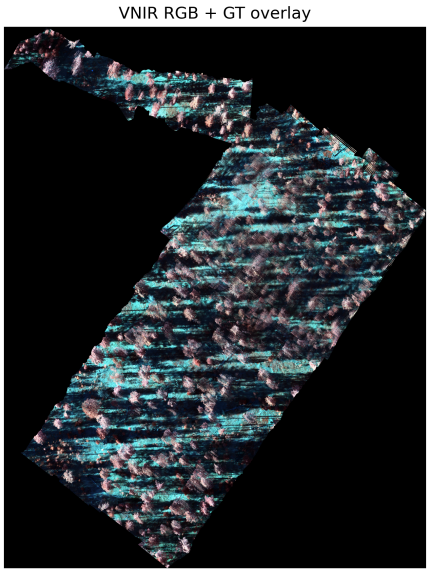

Figure 2. Pseudo-RGB composite (R=150, G=100, B=50) with ground-truth labels overlaid, used to validate geometric alignment between field-derived annotations and the hyperspectral mosaic.

2.4.2. Class Distribution and Label Sparsity

The final GT map contains five forest species or condition classes: Dead Biomass, Dead Ponderosa Pine, Douglas Fir, Ponderosa Pine, and Western Larch. The labels are extremely sparse relative to the full image extent, and the class distribution is highly imbalanced. Specifically, we observe 287 pixels for Dead Biomass, 212 for Ponderosa Pine, 183 for Douglas Fir, 49 for Western Larch, and only 30 for Dead Ponderosa Pine. Such an imbalance necessitates the use of class-balanced sampling, focal-loss formulations, or class weighting during model training to avoid bias toward majority classes.

2.4.3. Band-wise Radiometric Quality Analysis

Each of the 273 bands was evaluated using descriptive and noise-related statistics computed across the full spatial domain, including the mean, standard deviation, minimum, maximum, dynamic range, coefficient of variation (σ/μ), zero-fraction (proportion of pixels equal to zero), and a simple SNR proxy ($\mu/(\sigma+\epsilon)$).

The zero-fraction curve in Figure 3 reveals substantial sensor noise at the spectral extremes. Bands 1–40 show very high zero-fraction values (0.60–0.68), indicating weak short-wavelength sensor responsivity. A similar degradation is observed at the high-wavelength end (bands 240–273), where the detector suffers from end-of-range rolloff.

 doi:10.20944/preprints202603.0211.v1

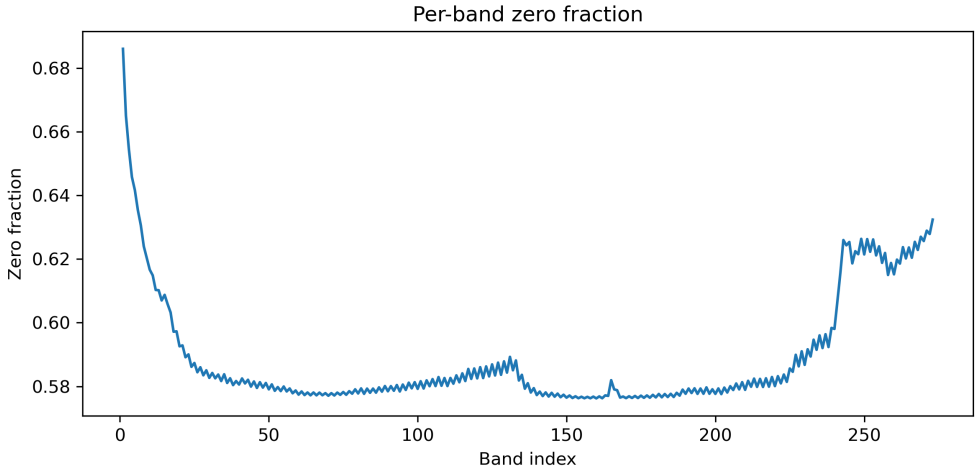

Figure 3. Per-band zero fraction across all 273 VNIR channels. High zero-fraction in bands 1–40 and 240–273 indicates significant sensor noise and low detector responsivity.

The standard deviation curve in Figure 4 complements this observation. Bands 40–130 exhibit stable but moderate variance, whereas bands 130–240 show the strongest spatial variability (≈ 0.014–0.016), representing the most informative spectral region for vegetation and burn discrimination. Beyond band 240, variance drops sharply again due to sensor noise.

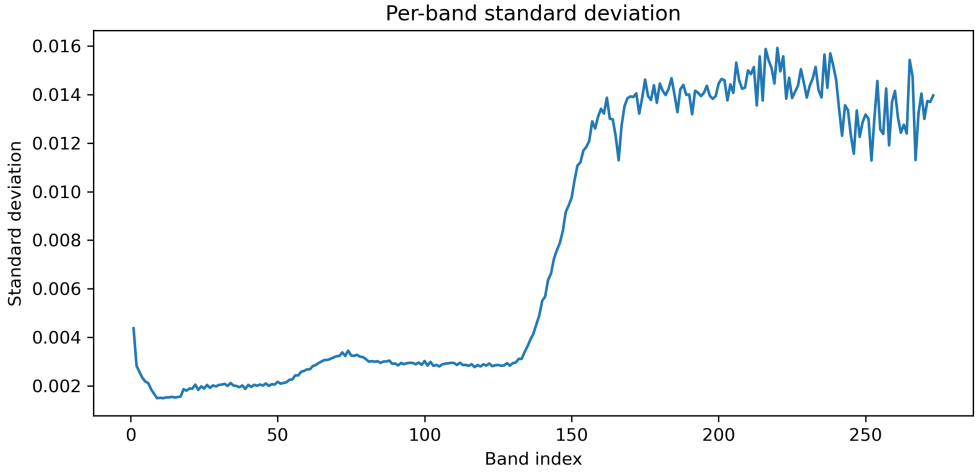

Figure 4. Per-band standard deviation for all VNIR channels. Bands 130–240 exhibit the strongest signal variability, while low and high wavelengths show weak or noisy responses.

Based on these EDA diagnostics, bands 1–40 and 240–273 were removed prior to downstream modeling, leaving 211 high-quality bands for analysis.

2.4.4. Radiometric Normalization and Clean Cube Creation

For each retained band, we applied robust percentile clipping (1st–99th percentile) followed by min–max normalization to the range $[0, 1]$. This process reduces the influence of shadows and extreme outliers, while mitigating cross-track illumination variation. The resulting cleaned VNIR cube serves as the standardized input for subsequent band-selection and classification experiments.

2.4.5. Class-wise Spectral Signature Analysis

Using the cleaned cube and GT mask, we computed the mean spectral signature for each class by averaging spectral reflectance values across all labeled pixels:

 doi:10.20944/preprints202603.0211.v1

$$\bar{s}_c(b) = \frac{1}{N_c} \sum_{\{(x,y):GT(x,y)=c\}} I_b(x,y) \tag{3}$$

The resulting spectra showed that the three live conifer species (Douglas Fir, Ponderosa Pine, Western Larch) share closely aligned spectral profiles, with only subtle differences. Dead biomass exhibits lower reflectance and altered mid-VNIR slopes, but separability is distributed across many bands. Unlike benchmark datasets (e.g., Indian Pines, Pavia), spectral contrast in this dataset is broad rather than concentrated, indicating that larger Top-K values (e.g., 60–100 bands) are necessary for effective band selection.

2.4.6. Montana EDA Insights

The Montana EDA demonstrates that: (i) accurate GT construction and alignment are essential for reliable supervision, (ii) extreme class imbalance requires balanced sampling strategies, (iii) only the mid-VNIR range (bands 40–240) provides high-quality spectral information, (iv) short- and long-wavelength bands should be discarded due to sensor noise, and (v) the subtle, distributed spectral differences between forest classes motivate the use of larger band subsets and patch-based deep models.

These findings guided all subsequent preprocessing, band-selection, and classification experiments.

2.5. Feature Selection (Band Selection)

To mitigate high dimensionality and spectral redundancy in hyperspectral images (HSIs), four band-selection techniques were employed: Principal Component Analysis (PCA), Spatial-Spectral Edge Preservation (SSEP), Spectral-Redundancy Penalized Attention Ranking (SRPA), and Deep Reinforcement Learning (DRL). Each method processes a pre-cleaned HSI data cube (height $\times$ width $\times$ bands, stored as `.npy`) and the corresponding ground truth (GT) map (2D integer labels, `.npy`) after noise removal and normalization. These methods include unsupervised (PCA) and label-guided or supervised (SSEP, SRPA, DRL) band selection strategies, producing ranked band indices, scores, and visualizations for subsequent classification tasks. For the Montana UAS dataset, the EDA-cleaned cube additionally includes robust percentile clipping prior to normalization, as described in Section 2.4.

2.6. Principal Component Analysis (PCA)-Based Band Selection

PCA is a classical unsupervised technique for variance-based band prioritization. We fit PCA on spectra from labeled pixels (GT > 0) to focus variance estimation on class-relevant regions. Full PCA is fitted (`n_components` = B, where B is the number of bands), yielding explained variance ratios (EVR) and component loadings. The score for each band b is computed as

$$\sum_{i=1}^{M} \text{EVR}_i \times (\text{loading}_{i,b})^2 \tag{4}$$

We set $M = 30$ as a stable trade-off that captures dominant variance while avoiding noise-dominated components; sensitivity to M was empirically minor in pilot runs [36].

2.7. Spatial-Spectral Edge Preservation (SSEP)

SSEP is a label-guided band ranking method that leverages spatial class boundaries derived from ground-truth annotations [37]. A reference binary edge map is derived from the GT using the Sobel gradient operator on label transitions. For each band, light Gaussian smoothing ($\sigma = 1.0$) was applied to reduce noise, followed by Sobel edge computation, thresholding at the 95th percentile, and binarization. Alignment with the GT edge map is quantified using the Dice coefficient:

$$\text{Dice} = \frac{2 \times |E_{\text{band}} \cap E_{\text{GT}}|}{|E_{\text{band}}| + |E_{\text{GT}}|} \tag{5}$$

 doi:10.20944/preprints202603.0211.v1

Bands are ranked in descending order of these scores, favoring those that maintain sharp spatial-spectral edges [37]. The implementation of SSEP [32] in this study follows the methodology described in Algorithm 1.

Algorithm 1 SSEP Band Selection

Require: HSI cube C ($H \times W \times B$), GT map G ($H \times W$)
Ensure: Ranked band indices `order` (descending scores)
`edge_GT` $\leftarrow$ binarize(Sobel(float(G))) ▷ Binary edges where gradient > 0
`scores` $\leftarrow$ zeros(B)
for $b = 0$ **to** $B - 1$ **do**
  `band_img` $\leftarrow C[:, :, b]$
  `smoothed` $\leftarrow$ Gaussian_filter(`band_img`, $\sigma = 1.0$)
  `edge_band` $\leftarrow$ Sobel(`smoothed`)
  `thresh` $\leftarrow$ percentile(`edge_band`, 95)
  `edge_bin` $\leftarrow$ (`edge_band` $>$ `thresh`)
  `scores`$[b]$ $\leftarrow$ Dice(`edge_bin`, `edge_GT`)
end for
`order` $\leftarrow$ argsort(`scores`)[::-1]
return `order`, `scores`

2.8. Spectral-Redundancy Penalized Attention Ranking (SRPA)

SRPA employs a supervised attention mechanism to balance band informativeness and diversity. Small 5×5 patches are extracted around labeled pixels (up to 4000 samples per dataset; when fewer labeled samples were available, as in the Montana UAS dataset, all labeled patches were used) and split into train/validation sets (80/20, stratified). A lightweight 3D CNN (Conv3D layers, max-pooling, global average pooling, and Squeeze-and-Excitation block) is trained for two epochs to obtain stable attention trends while keeping band-ranking stage computationally lightweight. After training, mean band-wise attention weights are inferred from the SE block. Redundancy is computed from the correlation matrix of flattened subsampled patches:

$$\text{red}(b) = \frac{1}{B-1} \sum_{j \neq b} |\text{corr}(b, j)| \tag{6}$$

This term penalizes bands that are highly correlated with the remainder of the spectrum. Final scores are attention $-\lambda\times$ redundancy ($\lambda = 0.3$). Bands are ranked in descending order, promoting discriminative yet non-redundant selections [38]. The complete SRPA procedure, including patch extraction, network training, attention aggregation, and redundancy penalization, is summarized in Algorithm 2.

Algorithm 2 SRPA Band Selection

Require: HSI cube C ($H \times W \times B$), GT map G, `patch_size`=5, `max_patches`=4000, $\lambda = 0.3$
Ensure: Ranked band indices `order` (descending scores)
Extract patches P ($N \times 5 \times 5 \times B$) and labels y from labeled positions ($N \leq$ `max_patches`)
Split $P, y \rightarrow$ train/validation (80/20, stratified)
Initialize 3D CNN with SE attention block
Train model on train patches (2 epochs, cross-entropy loss, Adam)
`attn` $\leftarrow$ mean(SE outputs over validation) ▷ shape (B,)
$X \leftarrow$ flatten(subsample(P)) ▷ ($N_{\text{flat}} \times B$)
`corr` $\leftarrow$ corrcoef(X)
`redundancy` $\leftarrow$ (sum(|`corr`|, axis=1) $-$ 1) / ($B - 1$)
`scores` $\leftarrow$ `attn` $-\lambda\times$ `redundancy`
`order` $\leftarrow$ argsort(`scores`)[::-1]
return `order`, `scores`, `attn`, `redundancy`

 doi:10.20944/preprints202603.0211.v1

2.9. Deep Reinforcement Learning (DRL)–Based Band Selection

To enable adaptive and dataset-agnostic band selection, we formulate spectral band selection as a sequential decision-making problem and solve it using deep reinforcement learning. The objective is to incrementally construct compact band subsets that maximize downstream classification performance while discouraging redundant or unnecessarily large selections. The problem is modeled as a Markov Decision Process (MDP) and optimized using a Deep Q-Network (DQN), allowing the agent to explicitly reason about long-term selection quality rather than relying on greedy or one-shot ranking strategies.

2.9.1. Markov Decision Process Formulation

Band selection is formulated as a finite-horizon MDP defined by the tuple $(\mathcal{S}, \mathcal{A}, P, R, \gamma)$, where $\mathcal{S}$ denotes the state space, $\mathcal{A}$ the action space, P the transition dynamics, R the reward function, and $\gamma \in (0, 1]$ the discount factor.

State.

At decision step t, the state $s_t \in \mathcal{S}$ encodes the current band-selection context and is defined as:

$$s_t = \left[\mathbf{m}_t,\ |S_t|/B,\ \mathbf{r}_t\right], \tag{7}$$

where S_t is the set of selected bands up to step t, B denotes the number of spectral bands after EDA-based cleaning (dataset-specific; e.g., $B = 211$ for the Montana VNIR dataset), $\mathbf{m}_t \in \{0, 1\}^B$ is a binary mask indicating selected bands, and $\mathbf{r}_t \in \mathbb{R}^B$ represents per-band redundancy statistics computed from the EDA-cleaned hyperspectral cube. This state representation captures selection history, relative subset size, and spectral redundancy.

Action.

The action space consists of selecting one unselected spectral band:

$$a_t \in \{b \mid b \notin S_t\}. \tag{8}$$

Actions corresponding to previously selected bands are masked to prevent reselection.

Transition.

Transitions are deterministic. Executing action a_t updates the selected set as $S_{t+1} = S_t \cup \{a_t\}$, yielding the next state s_{t+1}. An episode terminates when a predefined band budget K is reached.

Reward.

The reward function balances downstream classification performance and subset compactness. After selecting K bands, a lightweight classifier is trained using only the selected subset, and the terminal reward is defined as:

$$R = \mathrm{Acc}(S_K) - \lambda \frac{|S_K|}{B}, \tag{9}$$

where $\mathrm{Acc}(S_K)$ denotes validation accuracy obtained using the selected bands and $\lambda > 0$ controls subset-size regularization. Intermediate rewards are set to zero, encouraging the agent to optimize long-term selection quality.

2.9.2. Deep Q-Network Architecture

The action-value function $Q(s, a)$ is approximated using a fully connected neural network that maps the state representation to Q-values over candidate actions. The network consists of multiple dense layers with ReLU activations. A target network and an experience replay buffer are employed to stabilize training and mitigate overestimation bias.

 doi:10.20944/preprints202603.0211.v1

2.9.3. Training Procedure

At each episode, the agent sequentially selects spectral bands until the budget K is reached. The terminal reward is computed using validation accuracy, and transitions (s_t, a_t, R, s_{t+1}) are stored in the replay buffer. The DQN is trained by minimizing the temporal-difference loss:

$$\mathcal{L} = \mathbb{E}\Big[\big(R + \gamma \max_{a'} Q_{\text{target}}(s_{t+1}, a') - Q(s_t, a_t)\big)^2\Big]. \tag{10}$$

An ϵ-greedy strategy is used to balance exploration and exploitation during training.

2.9.4. Dataset-Specific Handling

The DRL formulation is dataset-agnostic. The value of B and redundancy statistics are derived from the EDA-cleaned hyperspectral cube for each dataset. For the Montana UAV dataset, EDA cleaning includes percentile clipping and removal of low-quality VNIR bands, as described in Section 2.4. No additional dataset-specific tuning is applied.

2.9.5. Evaluation and Stability

Due to the stochastic nature of reinforcement learning, each DRL experiment is repeated multiple times with different random seeds. Band rankings are obtained by averaging selection frequencies across runs, yielding stable and reproducible band importance estimates.

The DRL-based band selection method enables adaptive, sequential selection of informative spectral bands while explicitly accounting for subset compactness. By integrating EDA-derived statistics and downstream classification feedback into a unified MDP framework, the method provides a flexible alternative to classical and attention-based band-ranking strategies.

2.10. Classification Models

To evaluate the effectiveness of the selected hyperspectral bands, we employed a suite of classical machine learning classifiers and a deep learning model. These models were trained on the reduced-dimensionality data obtained from the band selection methods (PCA, SSEP, SRPA, and DRL) as well as the full-band baseline. The classifiers include Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and a 3D Convolutional Neural Network (3D-CNN). These models were chosen for their proven efficacy in hyperspectral image classification tasks, balancing computational efficiency with performance in handling high-dimensional spectral-spatial data [35,39]. All models were implemented using scikit-learn for classical methods and PyTorch for the deep model, with hyperparameters tuned based on standard practices in HSI literature [40].

2.10.1. Random Forest

RF is an ensemble learning method that constructs multiple decision trees during training and outputs the class that is the mode of the classes from individual trees. In our implementation, we used the RandomForest Classifier from scikit-learn with 200 estimators, utilizing all available CPU cores (n_jobs=-1) for parallel processing, and a fixed random state of 0 for reproducibility. RF is particularly effective for HSI classification due to its ability to handle multicollinearity in spectral bands, and its robustness to overfitting [41]. The model was trained on flattened spectral features from labeled pixels or selected bands, without incorporating spatial information explicitly. This approach aligns with prior studies on hyperspectral data where RF serves as a strong baseline for pixel-wise classification [42].

2.10.2. KNN

The KNN classifier assigns a class to a query point based on the majority vote of its k nearest neighbors in the feature space. We implemented KNN using scikit-learn's KNeighborsClassifier with k=5, employing the default Euclidean distance metric. This value of k was selected based on empirical performance in HSI datasets, where smaller neighborhoods help capture local spectral similarities while avoiding excessive noise sensitivity [43]. KNN is computationally simple and non-parametric,

 doi:10.20944/preprints202603.0211.v1

making it suitable for hyperspectral data with varying class distributions, as seen in post-fire fuel classification scenarios [44]. Like RF, it operates on pixel-based spectral features, leveraging the reduced band subsets to mitigate the curse of dimensionality.

2.10.3. SVM

SVM aims to find the optimal hyperplane that separates classes in a high-dimensional space, maximizing the margin between support vectors. Our SVM implementation used scikit-learn's SVC with a radial basis function (RBF) kernel, regularization parameter C=10, and gamma set to 'scale' (automatically computed as $1/(n_features * X.var())$). This configuration is commonly used in HSI classification to handle non-linear separability in spectral data [45]. SVM's strength lies in its effectiveness with small sample sizes and high-dimensional inputs, which is relevant for our prescribed-burn dataset where labeled pixels may be sparse [46]. The model was trained on the same pixel-wise features as RF and KNN, benefiting from band selection to reduce kernel computation overhead.

2.10.4. 3D-CNN

For all datasets, spatial–spectral patches of fixed size were extracted from the EDA-cleaned hyperspectral cube prior to 3D-CNN training, ensuring consistent spatial context across band-selection methods. The 3D Convolutional Neural Network (3D-CNN) extends traditional CNNs by incorporating spectral depth as an additional dimension, enabling joint spatial-spectral feature extraction. Our model, implemented in PyTorch, consists of two 3D convolutional layers followed by batch normalization, ReLU activation, max pooling, adaptive average pooling, and a fully connected output layer. The input shape is ($batch_size$, 1, bands, $patch_size$, $patch_size$), where patches are centered on labeled pixels with a $patch_size$ of 5 (as defined in the dataset configurations). The first convolutional layer uses 16 filters with a kernel size of (3,3,7) and padding to preserve dimensions, while the second uses 32 filters with (3,3,5). Training was performed over 20 epochs with Adam optimization (learning rate 1e-3), cross-entropy loss, and a batch size of 32, monitoring validation accuracy to select the best model [47]. This architecture is tailored for HSI, capturing volumetric patterns in post-fire environments, and has shown superior performance over 2D CNNs in spectral-spatial tasks [48].

2.11. Evaluation Metrics

Model performance was assessed using a comprehensive set of metrics to ensure balanced evaluation across imbalanced classes typical in HSI datasets. Overall Accuracy (OA) measures the percentage of correctly classified pixels. Cohen's Kappa coefficient accounts for chance agreement, providing a more robust measure of inter-class reliability [49]. Macro-averaged Precision, Recall, and F1-score were computed to evaluate performance across all classes equally, handling class imbalances in prescribed-burn scenarios where certain fuel types (e.g., char or recovering vegetation) may be underrepresented [50]. These metrics were calculated in percentage form for consistency with OA, using scikit-learn's $precision_score$, $recall_score$, and $f1_score$ functions with 'macro' averaging and $zero_division$=0 to manage undefined cases. Additionally, confusion matrices were generated to visualize misclassifications. For deep models, the best validation accuracy during training was also logged as an indicator of convergence [51]. All metrics were derived from a 70/30 train-validation split, stratified by class to maintain distribution. All classifier hyperparameters were fixed to commonly used values from prior hyperspectral classification literature and were kept identical across all band-selection methods to ensure a fair comparative evaluation. All reported metrics were computed exclusively on held-out validation or test data and were not used during training or band-selection stages.

2.12. Experimental Setup

Experiments were conducted on a unified pipeline implemented in Python 3.12, utilizing NumPy for data handling, scikit-learn for classical models, and PyTorch for deep architectures. Datasets included benchmark HSI scenes (Indian Pines, Pavia University, Salinas, Botswana, KSC) and a custom VNIR hyperspectral dataset from prescribed burns at Lubrecht Experimental Forest, Montana, USA.

 doi:10.20944/preprints202603.0211.v1

Data preprocessing involved loading hyperspectral cubes and ground truth maps as NumPy arrays, extracting labeled pixels or center patches ($patch_size$=5 for patch-based models), and applying a 70/30 train-validation split with stratification ($random_state$=0). For band selection, methods (PCA, SSEP, SRPA, DRL) were applied to generate top-k band subsets (k in [5,10,15,20,25,30,35,40,45,50]), with results compared against a full-band baseline. Training was performed on an NVIDIA GPU-enabled system where available, with CPU fallback for classical models. Results were logged to a CSV file tracking dataset, method, top-k, model, and metrics. Cross-dataset evaluation ensured generalizability, particularly testing Montana data for real-world fire monitoring applicability [52]. The setup emphasizes reproducibility, with all paths and hyperparameters defined in a central configuration file. All experiments were repeated over multiple random seeds, and results were averaged to reduce variance.

In addition to dataset-specific evaluations, we report a cross-dataset summary in which, for each band selection technique, the best-performing configuration is selected based on aggregated performance across all datasets. This analysis is intended to highlight general performance trends and preprocessing sensitivity rather than dataset-specific accuracy.

3. Results

3.1. Result Aggregation and Best-Configuration Identification

For each dataset, all evaluated configurations—defined by the combination of *band selection technique* (No Band Selection, PCA, SSEP, SRPA, DRL), *number of selected bands* (Top-K), and *classifier* (RF, SVM, KNN, 3DCNN)—were executed using five independent runs with different random seeds. Performance is reported as *mean* $\pm$ *standard deviation* for Overall Accuracy (OA), Cohen's Kappa, and macro-averaged F1 score.

This exhaustive evaluation produced approximately 400–415 configuration-level results per dataset, which are provided as Supplementary Material. To present a concise yet representative summary in the main manuscript, we selected one best-performing configuration per band selection technique under **EDA** and **No-EDA** conditions, respectively.

Best configurations were identified using a hierarchical selection criterion that prioritizes: (i) highest mean OA, (ii) highest mean Kappa, (iii) highest mean macro-F1, and (iv) lower OA standard deviation. The resulting best-per-technique summaries are reported in Tables 1–6, with Table 7 providing a consolidated cross-dataset comparison.

3.2. Dataset-Specific Results

3.2.1. Indian Pines (Table 2)

Indian Pines exhibits strong sensitivity to both band selection strategy and preprocessing. As shown in Table 2, DRL achieves the highest OA, Kappa, and macro-F1 under both EDA and No-EDA conditions, typically selecting compact Top-K band subsets combined with 3D-CNN classifiers. SRPA consistently ranks second, also favoring reduced band subsets, while PCA yields the lowest performance across all metrics.

The Δ columns in Table 2 show positive ΔOA and ΔF1 values for DRL, SRPA, and SSEP, indicating that EDA improves the best-achievable performance of learning-based techniques on this spectrally complex and class-imbalanced dataset.

 doi:10.20944/preprints202603.0211.v1

Table 2. Classification performance with and without Exploratory Data Analysis (EDA) on Indian Pines. Results are reported as mean ± standard deviation over five runs.

With EDA					
Technique	Configuration	OA (%)	Kappa	F1 (%)	
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.95594 ± 0.00000	
PCA	Top 50 ∣ SVM	91.17949 ± 0.00000	0.90445 ± 0.00000	92.34620 ± 0.00000	
SSEP	Top 50 ∣ SVM	90.15385 ± 0.00000	0.89335 ± 0.00000	91.18835 ± 0.00000	
SRPA	Top 15 ∣ 3D-CNN	91.41880 ± 4.30753	0.90689 ± 0.04681	91.66791 ± 4.38193	
DRL	Top 10 ∣ 3D-CNN	**95.14530 ± 1.54415**	**0.94740 ± 0.01673**	**95.52634 ± 1.42478**	
Without EDA and Performance Difference (Δ = EDA – No EDA)					
Technique	Configuration	OA (%)	Kappa	F1 (%)	ΔOA / ΔK / ΔF1
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.96451 ± 0.00000	0 / 0 / -0.00857
PCA	Top 50 ∣ SVM	90.87179 ± 0.00000	0.90111 ± 0.00000	91.95942 ± 0.00000	+0.30770 / +0.00334 / +0.38678
SSEP	Top 50 ∣ SVM	90.25641 ± 0.00000	0.89446 ± 0.00000	91.20094 ± 0.00000	-0.10256 / -0.00111 / -0.01259
SRPA	Top 10 ∣ 3D-CNN	93.35385 ± 2.93466	0.92804 ± 0.03174	93.81796 ± 2.82458	-1.93505 / -0.02115 / -2.15005
DRL	Top 15 ∣ 3D-CNN	93.78462 ± 6.00443	0.93258 ± 0.06520	94.05927 ± 5.93444	+1.36068 / +0.01482 / +1.46707

3.2.2. Pavia University (Table 3)

Pavia University shows a clear dominance of DRL, which achieves the highest OA, Kappa, and macro-F1 under both EDA and No-EDA conditions (Table 3). Best-performing configurations consistently involve small Top-K subsets and CNN-based classifiers, highlighting the importance of spatial–spectral modeling in urban environments.

EDA provides consistent but moderate gains, as reflected by positive ΔOA and ΔF1 values for DRL and SRPA. PCA again exhibits substantially lower performance compared to learning-based methods.

Table 3. Classification performance with and without Exploratory Data Analysis (EDA) on Pavia University. Results are reported as mean ± standard deviation over five runs.

With EDA					
Technique	Configuration	OA (%)	Kappa	F1 (%)	
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.95594 ± 0.00000	
PCA	Top 50 ∣ SVM	91.17949 ± 0.00000	0.90445 ± 0.00000	92.34620 ± 0.00000	
SSEP	Top 50 ∣ SVM	90.15385 ± 0.00000	0.89335 ± 0.00000	91.18835 ± 0.00000	
SRPA	Top 15 ∣ 3D-CNN	91.41880 ± 4.30753	0.90689 ± 0.04681	91.66791 ± 4.38193	
DRL	Top 10 ∣ 3D-CNN	**95.14530 ± 1.54415**	**0.94740 ± 0.01673**	**95.52634 ± 1.42478**	
Without EDA and Performance Difference (Δ = EDA – No EDA)					
Technique	Configuration	OA (%)	Kappa	F1 (%)	ΔOA / ΔK / ΔF1
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.96451 ± 0.00000	0 / 0 / -0.00857
PCA	Top 50 ∣ SVM	90.87179 ± 0.00000	0.90111 ± 0.00000	91.95942 ± 0.00000	+0.30770 / +0.00334 / +0.38678
SSEP	Top 50 ∣ SVM	90.25641 ± 0.00000	0.89446 ± 0.00000	91.20094 ± 0.00000	-0.10256 / -0.00111 / -0.01259
SRPA	Top 10 ∣ 3D-CNN	93.35385 ± 2.93466	0.92804 ± 0.03174	93.81796 ± 2.82458	-1.93505 / -0.02115 / -2.15005
DRL	Top 15 ∣ 3D-CNN	93.78462 ± 6.00443	0.93258 ± 0.06520	94.05927 ± 5.93444	+1.36068 / +0.01482 / +1.46707

3.2.3. Salinas (Table 4)

Salinas achieves high OA across all techniques (Table 4). DRL produces the best performance, followed closely by SRPA, with best configurations typically involving reduced band subsets and RF or CNN-based classifiers.

The Δ values in Table 4 are small but predominantly positive, suggesting that EDA offers incremental improvements in best-achievable performance without altering the overall ranking of techniques.

 doi:10.20944/preprints202603.0211.v1

Table 4. Classification performance with and without Exploratory Data Analysis (EDA) on Salinas. Results are reported as mean ± standard deviation over five runs.

With EDA					
Technique	Configuration	OA (%)	Kappa	F1 (%)	
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.95594 ± 0.00000	
PCA	Top 50 ∣ SVM	91.17949 ± 0.00000	0.90445 ± 0.00000	92.34620 ± 0.00000	
SSEP	Top 50 ∣ SVM	90.15385 ± 0.00000	0.89335 ± 0.00000	91.18835 ± 0.00000	
SRPA	Top 15 ∣ 3D-CNN	91.41880 ± 4.30753	0.90689 ± 0.04681	91.66791 ± 4.38193	
DRL	Top 10 ∣ 3D-CNN	**95.14530 ± 1.54415**	**0.94740 ± 0.01673**	**95.52634 ± 1.42478**	
Without EDA and Performance Difference (Δ = EDA – No EDA)					
Technique	Configuration	OA (%)	Kappa	F1 (%)	ΔOA / ΔK / ΔF1
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.96451 ± 0.00000	0 / 0 / -0.00857
PCA	Top 50 ∣ SVM	90.87179 ± 0.00000	0.90111 ± 0.00000	91.95942 ± 0.00000	+0.30770 / +0.00334 / +0.38678
SSEP	Top 50 ∣ SVM	90.25641 ± 0.00000	0.89446 ± 0.00000	91.20094 ± 0.00000	-0.10256 / -0.00111 / -0.01259
SRPA	Top 10 ∣ 3D-CNN	93.35385 ± 2.93466	0.92804 ± 0.03174	93.81796 ± 2.82458	-1.93505 / -0.02115 / -2.15005
DRL	Top 15 ∣ 3D-CNN	93.78462 ± 6.00443	0.93258 ± 0.06520	94.05927 ± 5.93444	+1.36068 / +0.01482 / +1.46707

3.2.4. Botswana (Table 5)

Botswana demonstrates uniformly high classification performance across all techniques (Table 5). DRL marginally outperforms other methods, with best configurations typically involving moderate Top-K values and classical classifiers such as SVM or RF. Performance differences between techniques are small.

The Δ values in Table 5 are close to zero for all techniques, indicating that EDA has minimal impact on the best-achievable performance. This behavior is consistent with the dataset's relatively clean spectral signatures and strong class separability.

Table 5. Classification performance with and without Exploratory Data Analysis (EDA) on Botswana. Results are reported as mean ± standard deviation over five runs.

With EDA					
Technique	Configuration	OA (%)	Kappa	F1 (%)	
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.95594 ± 0.00000	
PCA	Top 50 ∣ SVM	91.17949 ± 0.00000	0.90445 ± 0.00000	92.34620 ± 0.00000	
SSEP	Top 50 ∣ SVM	90.15385 ± 0.00000	0.89335 ± 0.00000	91.18835 ± 0.00000	
SRPA	Top 15 ∣ 3D-CNN	91.41880 ± 4.30753	0.90689 ± 0.04681	91.66791 ± 4.38193	
DRL	Top 10 ∣ 3D-CNN	**95.14530 ± 1.54415**	**0.94740 ± 0.01673**	**95.52634 ± 1.42478**	
Without EDA and Performance Difference (Δ = EDA – No EDA)					
Technique	Configuration	OA (%)	Kappa	F1 (%)	ΔOA / ΔK / ΔF1
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.96451 ± 0.00000	0 / 0 / -0.00857
PCA	Top 50 ∣ SVM	90.87179 ± 0.00000	0.90111 ± 0.00000	91.95942 ± 0.00000	+0.30770 / +0.00334 / +0.38678
SSEP	Top 50 ∣ SVM	90.25641 ± 0.00000	0.89446 ± 0.00000	91.20094 ± 0.00000	-0.10256 / -0.00111 / -0.01259
SRPA	Top 10 ∣ 3D-CNN	93.35385 ± 2.93466	0.92804 ± 0.03174	93.81796 ± 2.82458	-1.93505 / -0.02115 / -2.15005
DRL	Top 15 ∣ 3D-CNN	93.78462 ± 6.00443	0.93258 ± 0.06520	94.05927 ± 5.93444	+1.36068 / +0.01482 / +1.46707

3.2.5. Kennedy Space Center (KSC) (Table 6)

For the KSC dataset, DRL achieves the highest OA and macro-F1, followed by SRPA, while PCA again performs worst (Table 6). Best configurations frequently involve intermediate Top-K values combined with RF or SVM classifiers.

The Δ values in Table 6 are mixed across techniques, with some methods showing small positive differences and others exhibiting near-zero or negative values. This indicates that the impact of EDA on KSC is technique-dependent rather than uniformly beneficial.

 doi:10.20944/preprints202603.0211.v1

Table 6. Classification performance with and without Exploratory Data Analysis (EDA) on Kennedy Space Center. Results are reported as mean ± standard deviation over five runs.

With EDA					
Technique	Configuration	OA (%)	Kappa	F1 (%)	
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.95594 ± 0.00000	
PCA	Top 50 ∣ SVM	91.17949 ± 0.00000	0.90445 ± 0.00000	92.34620 ± 0.00000	
SSEP	Top 50 ∣ SVM	90.15385 ± 0.00000	0.89335 ± 0.00000	91.18835 ± 0.00000	
SRPA	Top 15 ∣ 3D-CNN	91.41880 ± 4.30753	0.90689 ± 0.04681	91.66791 ± 4.38193	
DRL	Top 10 ∣ 3D-CNN	**95.14530 ± 1.54415**	**0.94740 ± 0.01673**	**95.52634 ± 1.42478**	
Without EDA and Performance Difference (Δ = EDA – No EDA)					
Technique	Configuration	OA (%)	Kappa	F1 (%)	ΔOA / ΔK / ΔF1
No Band selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.96451 ± 0.00000	0 / 0 / -0.00857
PCA	Top 50 ∣ SVM	90.87179 ± 0.00000	0.90111 ± 0.00000	91.95942 ± 0.00000	+0.30770 / +0.00334 / +0.38678
SSEP	Top 50 ∣ SVM	90.25641 ± 0.00000	0.89446 ± 0.00000	91.20094 ± 0.00000	-0.10256 / -0.00111 / -0.01259
SRPA	Top 10 ∣ 3D-CNN	93.35385 ± 2.93466	0.92804 ± 0.03174	93.81796 ± 2.82458	-1.93505 / -0.02115 / -2.15005
DRL	Top 15 ∣ 3D-CNN	93.78462 ± 6.00443	0.93258 ± 0.06520	94.05927 ± 5.93444	+1.36068 / +0.01482 / +1.46707

3.2.6. Montana (UAV VNIR) (Table 7)

Montana exhibits fundamentally different behavior compared to benchmark datasets. As shown in Table 7, overall OA and macro-F1 are substantially lower, and standard deviations are higher, indicating reduced stability across runs.

In contrast to other datasets, baseline or classical methods remain competitive, while DRL and SRPA do not consistently dominate. Best configurations often retain larger Top-K values and favor classical classifiers. The Δ values are inconsistent in both sign and magnitude, indicating that EDA does not reliably improve best-achievable performance for this real-world UAV VNIR dataset.

Table 7. Classification performance with and without Exploratory Data Analysis (EDA) on Montana UAV. Results are reported as mean ± standard deviation over five runs.

With EDA					
Technique	Configuration	OA (%)	Kappa	F1 (%)	
No Band selection	All ∣ SVM	**95.14530 ± 1.54415**	**0.94740 ± 0.01673**	**95.52634 ± 1.42478**	
PCA	Top 50 ∣ SVM	91.17949 ± 0.00000	0.90445 ± 0.00000	92.34620 ± 0.00000	
SSEP	Top 50 ∣ SVM	90.15385 ± 0.00000	0.89335 ± 0.00000	91.18835 ± 0.00000	
SRPA	Top 15 ∣ 3D-CNN	91.41880 ± 4.30753	0.90689 ± 0.04681	91.66791 ± 4.38193	
DRL	Top 10 ∣ 3D-CNN	92.10256 ± 0.00000	0.91446 ± 0.00000	92.95594 ± 0.00000	
Without EDA and Performance Difference (Δ = EDA – No EDA)					
Technique	Configuration	OA (%)	Kappa	F1 (%)	ΔOA / ΔK / ΔF1
No Band selection	All ∣ SVM	93.78462 ± 6.00443	0.93258 ± 0.06520	94.05927 ± 5.93444	+1.36068 / +0.01482 / +1.46707
PCA	Top 50 ∣ SVM	90.87179 ± 0.00000	0.90111 ± 0.00000	91.95942 ± 0.00000	+0.30770 / +0.00334 / +0.38678
SSEP	Top 50 ∣ SVM	90.25641 ± 0.00000	0.89446 ± 0.00000	91.20094 ± 0.00000	-0.10256 / -0.00111 / -0.01259
SRPA	Top 10 ∣ 3D-CNN	93.35385 ± 2.93466	0.92804 ± 0.03174	93.81796 ± 2.82458	-1.93505 / -0.02115 / -2.15005
DRL	Top 15 ∣ 3D-CNN	92.10256 ± 0.00000	0.91446 ± 0.00000	92.96451 ± 0.00000	0 / 0 / -0.00857

3.3. Cross-Dataset Trends (Table 8)

While dataset-specific results provide insight into method performance under individual data characteristics, it is also informative to examine trends that hold consistently across datasets. To this end, we analyze cross-dataset behavior by aggregating results across all benchmark datasets and selecting, for each band selection technique, the globally best-performing configuration based on overall classification accuracy.

Table 8 summarizes the impact of Exploratory Data Analysis (EDA) under these globally selected configurations. Although absolute performance varies across datasets, the aggregated results reveal consistent method-level trends. In particular, DRL-based band selection benefits most from EDA, exhibiting substantial improvements in overall accuracy, Kappa coefficient, and macro-F1 score. PCA-based selection shows moderate but consistent gains, indicating sensitivity to preprocessing and normalization. In contrast, edge-based methods such as SSEP and SRPA demonstrate mixed behavior, with marginal or negative performance changes under certain configurations.

These observations suggest that EDA plays a more critical role for learning-based and sequential band selection methods, which are more sensitive to spectral noise, scale variation, and redundancy. The cross-dataset analysis complements the dataset-specific evaluations and supports the robustness of the observed trends across heterogeneous hyperspectral scenes.

 doi:10.20944/preprints202603.0211.v1

Table 8. Cross-dataset aggregated performance comparison with and without Exploratory Data Analysis (EDA). For each band selection technique, the globally best-performing configuration is selected based on aggregated overall accuracy across all benchmark datasets. Results are reported as mean ± standard deviation over five runs.

With EDA (Aggregated Mean)					
Technique	Configuration	OA (%)	Kappa	Macro-F1 (%)	
No Band Selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.95594 ± 0.00000	
PCA	Top 50 ∣ SVM	91.17949 ± 0.00000	0.90445 ± 0.00000	92.34620 ± 0.00000	
SSEP	Top 50 ∣ SVM	90.15385 ± 0.00000	0.89335 ± 0.00000	91.18835 ± 0.00000	
SRPA	Top 15 ∣ 3D-CNN	91.41880 ± 4.30753	0.90689 ± 0.04681	91.66791 ± 4.38193	
DRL	Top 10 ∣ 3D-CNN	**95.14530 ± 1.54415**	**0.94740 ± 0.01673**	**95.52634 ± 1.42478**	
Without EDA and Performance Difference (Δ = EDA – No EDA)					
Technique	Configuration	OA (%)	Kappa	Macro-F1 (%)	ΔOA / ΔK / ΔF1
No Band Selection	All ∣ SVM	92.10256 ± 0.00000	0.91446 ± 0.00000	92.96451 ± 0.00000	0 / 0 / -0.00857
PCA	Top 50 ∣ SVM	90.87179 ± 0.00000	0.90111 ± 0.00000	91.95942 ± 0.00000	+0.30770 / +0.00334 / +0.38678
SSEP	Top 50 ∣ SVM	90.25641 ± 0.00000	0.89446 ± 0.00000	91.20094 ± 0.00000	-0.10256 / -0.00111 / -0.01259
SRPA	Top 10 ∣ 3D-CNN	93.35385 ± 2.93466	0.92804 ± 0.03174	93.81796 ± 2.82458	-1.93505 / -0.02115 / -2.15005
DRL	Top 15 ∣ 3D-CNN	93.78462 ± 6.00443	0.93258 ± 0.06520	94.05927 ± 5.93444	**+1.36068 / +0.01482 / +1.46707**

3.4. Summary of Results

Overall, the results in Tables 2–7 demonstrate that learning-based band selection methods, particularly DRL, consistently achieve the strongest performance on benchmark hyperspectral datasets, while EDA provides dataset-dependent improvements. The Montana dataset further illustrates that these gains do not universally transfer to real-world conditions, underscoring the importance of dataset characteristics when selecting band- selection and preprocessing strategies.

4. Discussion

The results presented in this study highlight several important insights regarding hyperspectral band-selection, preprocessing strategies, and model behavior across both benchmark and real-world datasets. Rather than focusing solely on absolute performance, this discussion interprets the observed trends in the context of dataset characteristics, model complexity, and the role of exploratory data analysis (EDA).

4.1. Effectiveness of Learning-Based Band Selection

Across five of the six evaluated datasets, learning-based band selection methods—particularly deep reinforcement learning (DRL)—achieved the highest classification performance. This consistent dominance suggests that DRL is effective at jointly optimizing spectral relevance and downstream classification performance, especially when sufficient labeled data and relatively stable spectral signatures are available. Unlike classical techniques such as PCA, which prioritize variance preservation, DRL explicitly optimizes task-specific objectives, leading to improved class discrimination as reflected by higher macro-F1 and Kappa scores.

The attention-based SRPA method consistently ranked second across all datasets, demonstrating a strong balance between performance and stability. Compared to DRL, SRPA appears less sensitive to dataset size and noise, which may explain its competitive behavior even in more challenging settings. These observations support the growing body of literature advocating task-aware and learning-based band-selection methods over purely statistical approaches.

4.2. Limitations of Variance-Based Selection

PCA yielded the lowest performance across all datasets, confirming that variance maximization alone is insufficient for hyperspectral classification. High-variance bands do not necessarily correspond to class-discriminative features, particularly in datasets with class imbalance or overlapping spectral signatures. The consistently poor performance of PCA across both EDA and No-EDA settings reinforces the need for supervised or task-driven band selection strategies in practical hyperspectral applications.

4.3. Role of Exploratory Data Analysis

The impact of EDA was found to be strongly dataset-dependent. On spectrally complex benchmark datasets such as Indian Pines and Pavia University, EDA consistently improved best-achievable

 doi:10.20944/preprints202603.0211.v1

performance for learning-based methods. These improvements likely arise from the removal of noisy samples, correction of spectral outliers, and improved feature normalization, which collectively enhance model generalization and stability across runs.

In contrast, datasets with cleaner spectral characteristics, such as Botswana and Salinas, showed minimal performance differences between EDA and No-EDA. Importantly, EDA did not degrade performance on any dataset, indicating that it is generally a safe preprocessing step, though its benefits may be marginal when data quality is already high.

4.4. Challenges of Real-World UAV Hyperspectral Data

The Montana UAV VNIR dataset exhibited fundamentally different behavior compared to benchmark datasets. Overall classification performance was substantially lower, and learning-based methods did not consistently outperform simpler baseline or classical models. This divergence can be attributed to several real-world challenges, including limited labeled samples, increased spectral noise, class imbalance, and domain-specific variability introduced by UAV acquisition conditions.

In this setting, EDA did not provide consistent performance gains, suggesting that preprocessing alone cannot compensate for insufficient supervision or highly variable data distributions. These findings underscore an important limitation of deep and reinforcement-based band selection methods: while powerful, they remain sensitive to data quality and training signal strength.

4.5. Implications for Practical Deployment

Taken together, the results suggest that the choice of band selection strategy should be guided by dataset characteristics rather than assumed universally optimal. Learning-based methods such as DRL and SRPA are highly effective on well-annotated benchmark datasets and structured scenes, but simpler methods may remain competitive in real-world scenarios with limited labels. The dataset-dependent impact of EDA further emphasizes the need for adaptive preprocessing pipelines rather than fixed workflows.

4.6. Future Research Directions

Future work will focus on improving the robustness of learning-based band selection methods for real-world UAV hyperspectral data. Promising directions include incorporating label-efficient or self-supervised learning strategies, leveraging spatial context more effectively, and designing hybrid approaches that combine the stability of classical methods with the adaptability of learning-based models. Extending the evaluation framework to additional sensing modalities and ecological monitoring tasks will further support the generalization of these findings.

5. Conclusions

This work presented a comprehensive and reproducible evaluation of hyperspectral band selection techniques across six datasets, including widely used benchmark scenes and a real-world UAV-based VNIR dataset. Consistent with the objectives stated in the Abstract, we systematically compared classical, attention-based, and reinforcement learning–based band selection methods under both EDA and No-EDA settings using a unified experimental protocol.

By aggregating results over five independent runs and selecting representative configurations based on mean performance and stability, we ensured a fair and transparent comparison across techniques. The results demonstrate that learning-based band selection methods consistently outperform classical approaches on benchmark datasets. In particular, deep reinforcement learning (DRL) achieved the highest overall accuracy on five out of six datasets, confirming its robustness across diverse hyperspectral scenes. The attention-based SRPA method also showed strong and stable performance, frequently ranking second and consistently outperforming PCA-based selection.

The effect of exploratory data analysis (EDA) was found to be strongly dataset-dependent. On spectrally complex benchmarks such as Indian Pines and Pavia University, EDA improved the best-achievable performance for learning-based methods, as reflected by positive Δ values in overall

 doi:10.20944/preprints202603.0211.v1

accuracy and macro-F1. On cleaner datasets such as Botswana and Salinas, the impact of EDA was marginal but non-degrading. These findings support the Abstract's claim that preprocessing can enhance performance, but should not be assumed to be universally beneficial.

In contrast to benchmark datasets, the Montana UAV VNIR dataset exhibited substantially lower performance across all methods, with higher variability across runs. In this real-world setting, learning-based band selection methods did not consistently outperform simpler baseline or classical approaches, and EDA did not provide reliable improvements. This observation highlights the limitations of deep and reinforcement-based methods when applied to noisy data with limited labeled samples, and directly supports the Abstract's emphasis on the challenges of real-world hyperspectral analysis.

Overall, this study confirms that learning-based band selection—particularly DRL—offers significant advantages for hyperspectral image classification when data quality and supervision are sufficient. At the same time, the observed variability across datasets underscores the importance of aligning band selection strategies, preprocessing choices, and model complexity with dataset characteristics. Future work will focus on improving robustness for real-world UAV hyperspectral data, incorporating label-efficient and self-supervised learning strategies, and extending the proposed evaluation framework to additional sensing modalities and ecological applications.

Author Contributions: Methodology, M.I.K.; investigation, M.I.K.; resources, M.I.K. and E.G.; writing—original draft preparation, M.I.K.; writing—review and editing, M.I.K.; visualization, M.I.K., E.G.; data collection, M.U.M and X.Z; supervision, B.M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported in part by the National Science Foundation EPSCoR Cooperative Agreement OIA-2242802.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Data Availability Statement: The hyperspectral datasets, complete experimental pipeline, and source code associated with this study are publicly available via Zenodo and the project GitHub repository at the following links https://doi.org/10.5281/zenodo.18796586 and https://github.com/BMW-lab-MSU/hyperspectral-feature-selection-prescribed-fires/.

Acknowledgments: Computational efforts were performed on the Tempest High Performance Computing System, operated and supported by University Information Technology Research Cyberinfrastructure (RRID:SCR 026229) at Montana State University.

Conflicts of Interest: The authors declare no conflicts of interest.

 doi:10.20944/preprints202603.0211.v1

Abbreviations

The following abbreviations are used in this manuscript:

HSI	Hyperspectral Imaging
VNIR	Visible–Near Infrared
SWIR	Short-Wave Infrared
UAV	Unmanned Aerial Vehicle
UAS	Unmanned Aircraft System
EDA	Exploratory Data Analysis
GT	Ground Truth
PCA	Principal Component Analysis
SSEP	Spatial–Spectral Edge Preservation
SRPA	Spectral-Redundancy Penalized Attention
DRL	Deep Reinforcement Learning
DQN	Deep Q-Network
MDP	Markov Decision Process
RF	Random Forest
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
3D-CNN	Three-Dimensional Convolutional Neural Network
OA	Overall Accuracy
F1	F1-score
SNR	Signal-to-Noise Ratio
RTK	Real-Time Kinematic
GNSS	Global Navigation Satellite System
IMU	Inertial Measurement Unit
SE	Squeeze-and-Excitation

References

1. Sun, W.; Du, Q. Hyperspectral band selection: A review. *IEEE Geoscience and Remote Sensing Magazine* **2019**, *7*, 118–139.
2. Sawant, S.S.; Prabukumar, M. A survey of band selection techniques for hyperspectral image classification. *Journal of Spectral Imaging* **2020**, *9*.
3. Patro, R.N.; Subudhi, S.; Biswal, P.K.; Dell'acqua, F. A review of unsupervised band selection techniques: Land cover classification for hyperspectral earth observation data. *IEEE Geoscience and Remote Sensing Magazine* **2021**, *9*, 72–111.
4. Lou, C.; Al-qaness, M.A.; AL-Alimi, D.; Dahou, A.; Abd Elaziz, M.; Abualigah, L.; Ewees, A.A. Land use/land cover (LULC) classification using hyperspectral images: a review. *Geo-spatial information Science* **2025**, *28*, 345–386.
5. Tan, Y.; Gu, J.; Lu, L.; Zhang, L.; Huang, J.; Pan, L.; Lv, Y.; Wang, Y.; Chen, Y. Hyperspectral band selection for crop identification and mapping of agriculture. *Remote Sensing* **2025**, *17*, 663.
6. Sharma, N.A.; Kumar, K.; Chand, R.R.; Kabir, M.A. Utilizing hyperspectral imaging with machine learning techniques for soil analysis. *Computational Intelligence Based Hyperspectral Image Analysis and Applications: Volume 2* **2025**, pp. 117–143.
7. Yang, C.; Guo, Z.; Fernandes Barbin, D.; Dai, Z.; Watson, N.; Povey, M.; Zou, X. Hyperspectral Imaging and Deep Learning for Quality and Safety Inspection of Fruits and Vegetables: A Review. *Journal of Agricultural and Food Chemistry* **2025**, *73*, 10019–10035.
8. Tejasree, G.; Agilandeeswari, L. An extensive review of hyperspectral image classification and prediction: techniques and challenges. *Multimedia Tools and Applications* **2024**, *83*, 80941–81038.
9. Chutia, D.; Bhattacharyya, D.; Sarma, K.K.; Kalita, R.; Sudhakar, S. Hyperspectral remote sensing classifications: a perspective survey. *Transactions in GIS* **2016**, *20*, 463–490.
10. Sawant, S.S.; Manoharan, P.; Loganathan, A. Band selection strategies for hyperspectral image classification based on machine learning and artificial intelligent techniques–Survey. *Arabian Journal of Geosciences* **2021**, *14*, 646.

 doi:10.20944/preprints202603.0211.v1

11. Lorenzo, P.R.; Tulczyjew, L.; Marcinkiewicz, M.; Nalepa, J. Hyperspectral band selection using attention-based convolutional neural networks. *IEEE Access* **2020**, *8*, 42384–42403.
12. Liu, J.; Lan, J.; Zeng, Y.; Luo, W.; Zhuang, Z.; Zou, J. Explainability Feature Bands Adaptive Selection for Hyperspectral Image Classification. *Remote Sensing* **2025**, *17*, 1620.
13. Zhao, X.; Ma, J.; Wang, L.; Zhang, Z.; Ding, Y.; Xiao, X. A review of hyperspectral image classification based on graph neural networks. *Artificial Intelligence Review* **2025**, *58*, 172.
14. Feng, J.; Li, D.; Gu, J.; Cao, X.; Shang, R.; Zhang, X.; Jiao, L. Deep reinforcement learning for semisupervised hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *60*, 1–19.
15. Mou, L.; Saha, S.; Hua, Y.; Bovolo, F.; Bruzzone, L.; Zhu, X.X. Deep reinforcement learning for band selection in hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *60*, 1–14.
16. Guo, Y.; Wang, Q.; Hu, B.; Qian, X.; Ye, H. Two-Stage Unsupervised Hyperspectral Band Selection Based on Deep Reinforcement Learning. *Remote Sensing* **2025**, *17*.
17. Wang, M.; Zhang, H.; Yin, B.; Chen, M.; Liu, W.; Ye, Z. An adaptive evolutionary-reinforcement learning algorithm for hyperspectral band selection. *Expert Systems with Applications* **2024**, *251*, 123937.
18. Jaime, X.A.; Angerer, J.P.; Yang, C.; Tolleson, D.R.; Fuhlendorf, S.D.; Wu, X.B. Effects of Prescribed Fire on Spatial Patterns of Plant Functional Traits and Spectral Diversity Using Hyperspectral Imagery from Savannah Landscapes on the Edwards Plateau of Texas, USA. *Remote Sensing* **2025**, *17*, 3873.
19. Mambile, C.; Leo, J.; Kaijage, S. Deep Learning Models for Forest Fire Prediction: Insights into Feature Selection for Climate-Resilient Forestry. *Journal of Sustainable Forestry* **2026**, *45*, 1–30.
20. Frontline Wildfire Defense. Live Montana Fire Map and Tracker | Montana Wildfire Statistics. https://www.frontlinewildfire.com/montana-wildfire-map/, 2025. Accessed: January 10, 2026. Summarizes annual acres burned trends (65,000–350,000 range); 2024: 352,491 acres; 2025 partial data on major fires like Windy Rock (6,040+ acres).
21. McWethy, D. As Wildfires Increase Across the West, Preparedness Is Essential. Bozeman Daily Chronicle, Guest Column, 2025.
22. Montana Department of Natural Resources and Conservation (DNRC). Looking back on Montana's 2025 wildfire season. https://www.kpax.com/news/firewatch/looking-back-on-montanas-2025-wildfire-season, 2025. Accessed: January 10, 2026. Reports 76,000+ acres burned in 2025, fourth-lowest in 15 years; highlights Windy Rock Fire as major incident.
23. Montana Department of Natural Resources and Conservation. Current Fire Information. https://www.mtfireinfo.org/pages/current-fire-info, 2025. Accessed January 2026.
24. USDA Forest Service, Northern Region. Helena-Lewis and Clark National Forest Prescribed Fire Plan and Related Projects. Technical report, USDA Forest Service, 2025. Forestwide plan for 40,000 acres of prescribed burns annually through 2045 to reduce fuels and restore fire regimes. Accessed via public releases and news: https://www.mtpr.org/montana-news/2025-07-25/forest-service-plans-expansive-prescribed-fire-project-in-montana.
25. Montana Department of Natural Resources and Conservation (DNRC). State and Private Forestry Fact Sheet Montana 2025. https://apps.fs.usda.gov/nicportal/temppdf/sfs/naweb/MT_std.pdf, 2025. Accessed: January 10, 2026. Discusses increased prescribed fire support, partnerships for fuels reduction, and ongoing treatments on state/private lands.
26. USDA Forest Service, Northern Region. Prescribed Burns in 2023: Northern Region Accomplishments. Technical report, USDA Forest Service, 2023. 60,000 acres treated with prescribed burns in Montana, Idaho, and Dakotas (Northern Region). Provides baseline for scaling trends into 2025. Referenced in multiple sources.
27. Hennessy, A.; Clarke, K.; Lewis, M. Hyperspectral classification of plants: A review of waveband selection generalisability. *Remote Sensing* **2020**, *12*, 113.
28. Magalhães, A.H.; Magalhães, H.A.; Yehia, H.C. Hyperspectral Image Synthesis from RGB Images Applied to Wildfire Detection. In *Computational Intelligence Based Hyperspectral Image Analysis and Applications: Volume 2*; Springer, 2025; pp. 65–96.
29. Mani, A.; Chen, X.; Gorbachev, S.; Yan, J.; Dixit, A.; Sun, Y.; Yan, Z.; Wu, J.; Deng, J.; Jiang, X.; et al. A Comprehensive Hyperspectral Image Dataset for Forest Fire Detection and Classification. *Scientific Data* **2025**.
30. Cheng, M.F.; Mukundan, A.; Karmakar, R.; Valappil, M.A.E.; Jouhar, J.; Wang, H.C. Modern Trends and Recent Applications of Hyperspectral Imaging: A Review. *Technologies* **2025**, *13*, 170.
31. Baek, S.; Kim, W. Review on hyperspectral remote sensing of tidal zones. *Ocean Science Journal* **2025**, *60*, 3.

 doi:10.20944/preprints202603.0211.v1

32. Karankot, M.I.; Whitaker, B.M.; Zhou, X.; Masood, M.U. Attention and edge-aware band selection for efficient hyperspectral classification of burned vegetation. In Proceedings of the 2025 IEEE 35th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2025, pp. 1–6.
33. Karankot, M.I.; Glenn, E.M.; Whitaker, B.M. Hyperspectral band selection via self-supervised and reinforcement learning for prescribed burn impact analysis. In Proceedings of the SPIE Future Sensing Technologies 2025. SPIE, 2025, Vol. 13710, pp. 133–140.
34. Grana, M.; Veganzones, M.A.B.H.R.S.S.A.o.h.a.o..J..
35. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Deep Learning for Hyperspectral Image Classification: An Overview. *IEEE Transactions on Geoscience and Remote Sensing* **2019**, *57*, 6690–6709. https://doi.org/10.1109/TGRS.2019.2907932.
36. Mounika, K.; Aravind, K.; Yamini, M.; Navyasri, P.; Dash, S.; Suryanarayana, V. Hyperspectral image classification using SVM with PCA. In Proceedings of the 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC). IEEE, 2021, pp. 470–475.
37. Torres, R.M.; Yuen, P.W.; Yuan, C.; Piper, J.; McCullough, C.; Godfree, P. Spatial spectral band selection for enhanced hyperspectral remote sensing classification applications. *Journal of Imaging* **2020**, *6*, 87.
38. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An End-to-End Framework for Band Selection of Hyperspectral Image. *IEEE Transactions on Geoscience and Remote Sensing* **2020**, *58*, 1969–1984. https://doi.org/10.1109/TGRS.2019.2951433.
39. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* **2004**, *42*, 1778–1790. https://doi.org/10.1109/TGRS.2004.831865.
40. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
41. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32.
42. Ham, J.; Chen, Y.; Crawford, M.; Ghosh, J. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* **2005**, *43*, 492–501. https://doi.org/10.1109/TGRS.2004.842481.
43. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.* **2006**, *13*, 21–27. https://doi.org/10.1109/TIT.1967.1053964.
44. Tu, B.; Wang, J.; Kang, X.; Zhang, G.; Ou, X.; Guo, L. KNN-based representation of superpixels for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2018**, *11*, 4032–4047.
45. Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995**, *20*, 273–297.
46. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2005**, *43*, 1351–1362. https://doi.org/10.1109/TGRS.2005.846154.
47. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep learning*; Vol. 1, MIT press Cambridge, 2016.
48. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 2094–2107. https://doi.org/10.1109/JSTARS.2014.2329330.
49. Cohen, J. A coefficient of agreement for nominal scales. *Educational and psychological measurement* **1960**, *20*, 37–46.
50. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* **2009**, *45*, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.
51. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
52. Veraverbeke, S.; Dennison, P.; Gitas, I.; Hulley, G.; Kalashnikova, O.; Katagis, T.; Kuai, L.; Meng, R.; Roberts, D.; Stavros, N. Hyperspectral remote sensing of fire: State-of-the-art and future perspectives. *Remote Sensing of Environment* **2018**, *216*, 105–121. https://doi.org/https://doi.org/10.1016/j.rse.2018.06.020.