# PREDICTION AS FEEDBACK: CLOSED-LOOP DEEP UNFOLDING FOR VISUAL CLASSIFICATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Deep unfolding has emerged as a promising paradigm for visual classification by integrating model-based optimization with deep representation learning. However, existing unfolding-based classifiers often retain an open-loop design without an explicit feedback path from intermediate predictions to subsequent solver updates. To bridge this gap, we propose **Pre2Inf**, a closed-loop deep unfolding framework that uses prediction as feedback to guide subsequent sparse inference, enabling bidirectional interaction between sparse inference and class prediction. Specifically, we formulate a class-aware sparse inference model and unfold its half-quadratic splitting solver into sequential stages. Each stage incorporates tailored modules for the resulting subproblems, with intermediate predictions mapped to sample-adaptive relevance weights that modulate atom-wise regularization in subsequent coefficient updates. Moreover, we analyze the stability of prediction-guided coefficient updates by characterizing their sensitivity to relevance perturbations. Experiments on eight pixel-level and two image-level classification benchmarks demonstrate consistent improvements over existing unfolding methods. Ablation studies further show that prediction feedback improves class separability and reduces reconstruction ambiguity during inference.

## 1 INTRODUCTION

Visual classification aims to assign semantic labels to visual observations at different spatial granularities, from entire images to individual pixels. Accurate classification requires representations that capture class-relevant information while suppressing nuisance variation. Data-driven approaches, such as deep neural networks, learn expressive representations (He et al., 2016; Dosovitskiy et al., 2021), although their intermediate computations often lack an explicit model-based interpretation. Model-driven approaches, such as sparse coding, structure inference through explicit objectives and priors, but their effectiveness depends on the suitability of the modeling assumptions (Mairal et al., 2012; Chen et al., 2011). Deep unfolding bridges these perspectives by mapping iterative inference algorithms to trainable network stages, combining model-based structure with data-driven representation learning (Gregor & LeCun, 2010; Hershey et al., 2014; Monga et al., 2021).

Building on this paradigm, deep unfolding has been explored in natural-image classification (Yu et al., 2024) and hyperspectral image classification (HSIC) (Nie et al., 2025). In these applications, an unfolded solver can iteratively refine representations, which are then mapped to class predictions by a classification head. Without feedback from these predictions to subsequent solver updates, however, the inference process remains open-loop even when the solver and classifier are jointly optimized with classification losses. In sparse inference, this limitation be-

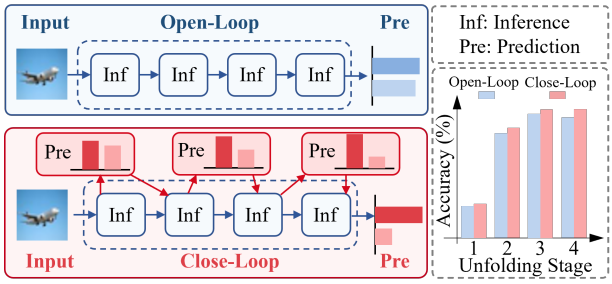



Figure 1: Motivation of closed-loop unfolding. The closed-loop method feeds intermediate predictions back to guide subsequent inference.



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

comes evident when multiple class-specific representations achieve similar reconstruction fidelity but differ in discriminative relevance (Mairal et al., 2012). This motivates a closed-loop formulation in which intermediate class evidence guides subsequent sparse inference toward more discriminative representations. As illustrated in Figure 1, closing this loop allows the solver to use evolving class estimates to resolve class-specific representational ambiguity: the solver improves the prediction, and the prediction, in turn, guides the solver.

To realize this interaction, we propose Pre2Inf, which formulates a class-aware sparse representation model and unfolds its half-quadratic splitting (HQS) solver into trainable stages. The key idea is to close the inference loop by feeding intermediate class predictions back into subsequent coefficient estimation. These predictions are converted into sample-adaptive relevance weights to modulate atom-wise regularization, while a confidence-aware mechanism controls their influence during iterative inference. Our main contributions are summarized as follows:

- We propose **Pre2Inf**, a closed-loop deep unfolding framework for visual classification that enables intermediate class predictions to participate in subsequent sparse inference. By coupling prediction and inference, Pre2Inf allows the optimization trajectory to adapt to evolving class evidence and resolve ambiguous class-specific representations.
- We formulate a class-aware sparse inference model and derive an HQS-based unfolding network with prediction-conditioned atom-wise regularization. The stability of feedback-conditioned coefficient updates is further analyzed by characterizing their well-posedness and sensitivity to relevance perturbations.
- Extensive experiments on eight hyperspectral benchmarks for pixel-level classification and CIFAR-10/100 for image-level classification (ImS) demonstrate the effectiveness and generality of Pre2Inf. Comprehensive ablation studies and stage-wise analyses further reveal how prediction feedback reshapes intermediate representations during iterative inference.

## 2 Related Work

**Deep unfolding for adaptive iterative inference.** Deep unfolding integrates model-based optimization with deep learning by transforming iterative algorithms into trainable architectures (Gregor & LeCun, 2010; Hershey et al., 2014; Monga et al., 2021). The learned iterative shrinkage and thresholding algorithm (LISTA) pioneered deep unfolding by unfolding sparse coding iterations into trainable networks (Gregor & LeCun, 2010). Subsequent studies enhanced unfolded inference with improved parameterization and adaptability, including ALISTA with analytical transformations (Liu et al., 2019), GLISTA with learnable gating (Wu et al., 2020), and ISTA-Net with learned nonlinear proximal mappings (Zhang & Ghanem, 2018). Recent works further incorporate powerful generative priors, such as diffusion models, into unfolding frameworks (Liao et al., 2025), while NSR-Net explores unfolding-based representation learning for hyperspectral classification (Nie et al., 2025). Despite these advances, existing methods mainly focus on improving solver components, such as update rules or learned priors, while the predictions produced during inference are generally treated as outputs rather than active signals for regulating subsequent optimization.

**Task-aware sparse inference and feedback-driven optimization.** Sparse representation methods have progressively incorporated task information to improve discriminative inference. Early approaches exploit class-specific reconstruction errors or representation structures for recognition (Chen et al., 2011), while task-driven dictionary learning and LC-KSVD introduce classification objectives or label consistency into representation learning (Mairal et al., 2012; Jiang et al., 2013). These methods demonstrate the importance of semantic information in sparse inference, but task cues are mainly encoded through learned models or predefined objectives. A more dynamic use of task information has been explored through feedback mechanisms, where previous predictions or intermediate states guide subsequent refinement, such as Iterative Error Feedback, Feedback Networks, and Recurrent Inference Machines (Carreira et al., 2016; Zamir et al., 2017; Putzky & Welling, 2017). Recent unfolding methods further incorporate auxiliary guidance during iterative inference (Dai et al., 2023; He et al., 2025; Alhejaili et al., 2025). However, these feedback mechanisms mainly provide guidance through representations or hidden states, without explicitly using task predictions to regulate the underlying optimization variables. Our work establishes such a coupling by using intermediate class predictions to regulate subsequent coefficient updates.

# 3 CLASS-GUIDED OPTIMIZATION MODEL AND ITERATIVE SOLUTION

We formulate classification as a closed-loop sparse inference process, where intermediate predictions regulate subsequent optimization. The resulting HQS solver enables prediction-guided coefficient refinement via adaptive atom-wise regularization.

## 3.1 MODEL FORMULATION

Given an input patch $\mathbf{Y} \in \mathbb{R}^{C\times H\times W}$, we consider the sparse observation model

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{E}, \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{C\times K}$ is a dictionary, $\mathbf{X} \in \mathbb{R}^{K\times H\times W}$ contains the sparse coefficients, and $\mathbf{E} \in \mathbb{R}^{C\times H\times W}$ captures residual corruption and modeling errors. Here, $H$, $W$, and $C$ denote the spatial height, width, and channel dimension of the input patch, respectively. The multiplication $\mathbf{DX}$ is performed independently at each spatial location, i.e., $(\mathbf{DX})_{:,h,w} = \mathbf{D}\mathbf{X}_{:,h,w}$. A reconstruction-oriented formulation estimates $\mathbf{X}$ and $\mathbf{E}$ by

$$\min_{\mathbf{X},\mathbf{E}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 + \lambda_x\|\mathbf{X}\|_1 + \lambda_e\|\mathbf{E}\|_{2,1}, \tag{2}$$

where $\lambda_x, \lambda_e \geq 0$ control the regularization strengths. The element-wise $\ell_1$ norm is defined as $\|\mathbf{X}\|_1 = \sum_{k,h,w} |X_{k,h,w}|$, and the group sparsity regularizer is defined as $\|\mathbf{E}\|_{2,1} = \sum_{h=1}^{H}\sum_{w=1}^{W} \|\mathbf{E}_{:,h,w}\|_2$. The two regularizers promote sparse coefficients and group-sparse residuals across spatial locations, respectively.

Existing unfolded classification frameworks generally adopt an open-loop inference process, where intermediate predictions are used only for supervision and do not affect subsequent coefficient updates. We incorporate this evidence through class-dependent coefficient regularization. The dictionary is partitioned according to $N$ semantic classes as $\mathbf{D} = [\mathbf{D}_1, \ldots, \mathbf{D}_N]$, where each class-specific sub-dictionary contains $m$ atoms, yielding $K = Nm$ total atoms. At stage $t \geq 1$, the current reconstruction $\mathbf{DX}^t$ is fed into a shared classifier to obtain class probabilities:

$$\mathbf{p}^t = \mathrm{softmax}\left(\frac{f_{\theta_{sg}}(\mathbf{D}\mathbf{X}^t)}{T_g}\right), \tag{3}$$

where $\mathbf{p}^t = [p_1^t, p_2^t, ..., p_N^t] \in \mathbb{R}^N$ denotes the predicted class distribution, $T_g$ is the guidance temperature, and $\theta_{sg}$ denotes stop-gradient. The predicted class evidence is then mapped to atom-wise relevance weights and combined with reconstruction-based similarity cues by $\mathcal{G}_\theta(.)$:

$$\boldsymbol{\alpha}^t = \mathcal{G}_\theta\left(\mathbf{p}^t, \mathbf{S}^t\right). \tag{4}$$

Here, $\mathbf{S}^t$ denotes the reconstruction–atom similarity and $\boldsymbol{\alpha}^t \in [0,1]^K$ provides class-conditioned regularization for subsequent coefficient updates. Conditioned on $\boldsymbol{\alpha}^t$, we formulate class-guided sparse inference as

$$\begin{aligned}\min_{\mathbf{X},\mathbf{E}} \quad & \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 + \lambda_x\|\mathbf{X}\|_1 + \lambda_e\|\mathbf{E}\|_{2,1} \\ & + \frac{\gamma}{2}\|(\mathbf{1} - \boldsymbol{\alpha}^t)^{1/2} \odot \mathbf{X}\|_F^2,\end{aligned} \tag{5}$$

where $\gamma \geq 0$ controls the strength of the prediction-conditioned regularization. The additional quadratic term assigns atom-wise penalties according to $\boldsymbol{\alpha}^t$, allowing class evidence from the current reconstruction to dynamically adapt the subsequent inference trajectory.

## 3.2 MODEL OPTIMIZATION

We apply HQS to decouple the sparse regularization from coefficient estimation and obtain stage-wise updates suitable for unfolding. Introducing an auxiliary variable $\mathbf{Z}$ and relaxing the constraint $\mathbf{Z} = \mathbf{X}$ with a quadratic penalty yields

$$\begin{aligned}\min_{\mathbf{X},\mathbf{Z},\mathbf{E}} \quad & \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X} - \mathbf{E}\|_F^2 + \lambda_x\|\mathbf{Z}\|_1 + \lambda_e\|\mathbf{E}\|_{2,1} \\ & + \frac{\gamma}{2}\|(\mathbf{1} - \boldsymbol{\alpha}^t)^{1/2} \odot \mathbf{X}\|_F^2 + \frac{\eta}{2}\|\mathbf{X} - \mathbf{Z}\|_F^2,\end{aligned} \tag{6}$$

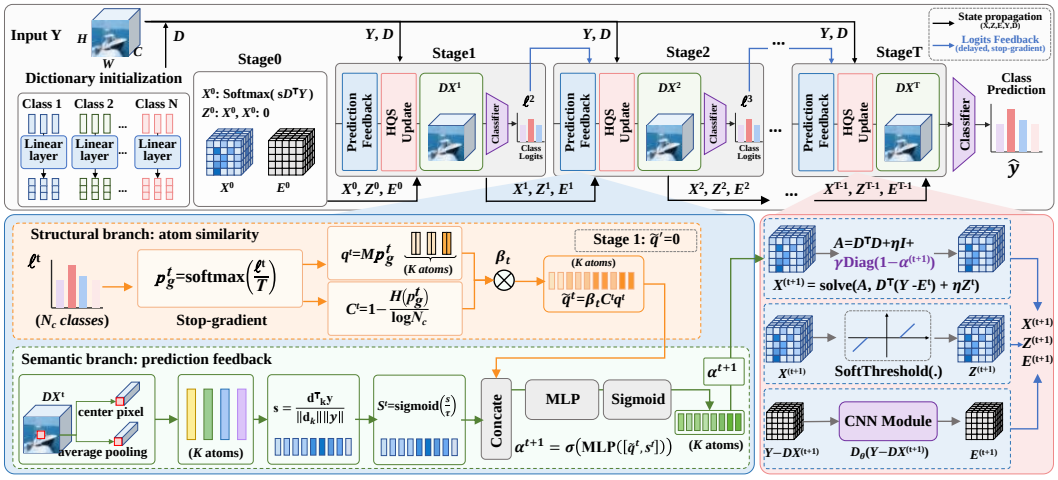



Figure 2: Overview of the proposed Pre2Inf framework. The framework unfolds HQS-based sparse inference into sequential stages, where intermediate predictions are fed back and transformed into adaptive atom-relevance weights to guide subsequent coefficient updates.

where $\eta > 0$. For fixed $\boldsymbol{\alpha}^t$, each unfolding stage performs one alternating update of $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{E}$.

$$\mathbf{A}^t\mathbf{X}^{t+1} = \mathbf{D}^\top(\mathbf{Y} - \mathbf{E}^t) + \eta\mathbf{Z}^t, \tag{7}$$

and

$$\mathbf{A}^t = \mathbf{D}^\top\mathbf{D} + \eta\mathbf{I} + \gamma\,\mathrm{Diag}(\mathbf{1} - \boldsymbol{\alpha}^t). \tag{8}$$

Since $\mathbf{A}^t \succeq \eta\mathbf{I}$, the coefficient subproblem admits a unique solution. Prediction feedback enters this system through its diagonal regularization, directly modifying the curvature of the coefficient subproblem.

Minimization with respect to $\mathbf{Z}$ gives element-wise soft-thresholding update, $\mathbf{Z}^{t+1} = \mathcal{S}_{\lambda_x/\eta}(\mathbf{X}^{t+1})$, where $\mathcal{S}_\tau(x) = \mathrm{sign}(x)\max(|x| - \tau, 0)$ denotes the element-wise soft-thresholding operator. For the residual $\mathbf{R}^{t+1} = \mathbf{Y} - \mathbf{D}\mathbf{X}^{t+1}$, the subproblem is

$$\min_{\mathbf{E}} \frac{1}{2}\|\mathbf{R}^{t+1} - \mathbf{E}\|_F^2 + \lambda_e\|\mathbf{E}\|_{2,1}. \tag{9}$$

In the unfolded network, we parameterize this residual refinement by a plug-and-play denoiser $\mathcal{M}_{\theta_e}$ while preserving the alternating update structure, without requiring it to be an exact proximal operator of the explicit $\ell_{2,1}$ penalty. The resulting updates are

$$\mathbf{X}^{t+1} = (\mathbf{A}^t)^{-1}\left[\mathbf{D}^\top(\mathbf{Y} - \mathbf{E}^t) + \eta\mathbf{Z}^t\right], \tag{10a}$$

$$\mathbf{Z}^{t+1} = \mathcal{S}_{\lambda_x/\eta}(\mathbf{X}^{t+1}), \tag{10b}$$

$$\mathbf{E}^{t+1} = \mathcal{M}_{\theta_e}(\mathbf{Y} - \mathbf{D}\mathbf{X}^{t+1}). \tag{10c}$$

The reconstruction $\mathbf{D}\mathbf{X}^{t+1}$ is then fed back to generate the next class estimate and relevance weights $\boldsymbol{\alpha}^{t+1}$. This iterative dependency couples prediction with sparse inference, forming the basis of the proposed closed-loop unfolding network and allowing the stability of prediction-conditioned coefficient updates to be analyzed under relevance perturbations.

## 4 Closed-Loop Class-Guided Deep Unfolding Network

### 4.1 Overall Framework

Based on the proposed class-aware sparse inference model, we unfold the iterative optimization process into a closed-loop deep unfolding network, as illustrated in Figure 2. Given an input hyperspectral patch $\mathbf{Y}$ and its dictionary $\mathbf{D}$, the network first employs a learnable initialization module to

obtain the initial sparse coefficient $\mathbf{X}^0$. The initialized variables are then passed through $T$ sequential unfolding stages, where each stage corresponds to one iteration of the proposed HQS solver. Specifically, the $t$-th unfolding stage updates the intermediate variables as

$$(\mathbf{X}^{t+1}, \mathbf{Z}^{t+1}, \mathbf{E}^{t+1}) = \mathcal{H}_\theta(\mathbf{X}^t, \mathbf{Z}^t, \mathbf{E}^t, \boldsymbol{\alpha}^t, \mathbf{Y}, \mathbf{D}), \tag{11}$$

where $\mathcal{H}_\theta$ denotes the learnable unfolding stage. Each stage updates the intermediate variables through the HQS solver and obtains the next guidance variable $\boldsymbol{\alpha}^{t+1}$ from the newly reconstructed representation $\mathbf{D}\mathbf{X}^{t+1}$. After $T$ unfolding stages, the prediction from the final unfolding stage is used for classification, and the network is optimized by jointly considering classification supervision and sparse inference consistency.

### 4.2 Learnable Dictionary-aware Initialization

The initialization provides the starting point for the subsequent unfolding process. Instead of using a fixed dictionary projection, we introduce a learnable dictionary-aware initialization module to obtain an adaptive initial coefficient estimate from the input patch and dictionary. Given an input patch $\mathbf{Y} \in \mathbb{R}^{C\times H\times W}$ and a class-structured dictionary $\mathbf{D} = [\mathbf{D}_1, \ldots, \mathbf{D}_N]$, we first transform the input and dictionary atoms into a learnable embedding space:

$$\tilde{\mathbf{Y}} = \phi_y(\mathbf{Y}), \qquad \tilde{\mathbf{D}} = \phi_d(\mathbf{D}), \tag{12}$$

where $\phi_y(\cdot)$ and $\phi_d(\cdot)$ denote learnable projection operations. These transformations preserve the class structure of the dictionary while adapting its atom representation, where the number of classes remains unchanged and each class contains $m'$ atoms, yielding $K = Nm'$ total atoms. The initial coefficient estimation is obtained from the correlation between the projected patch features and dictionary atoms, and the residual is initialized as zero:

$$\mathbf{X}^0 = \mathrm{Softmax}\left(s\tilde{\mathbf{D}}^\top\tilde{\mathbf{Y}}\right), \qquad \mathbf{E}^0 = \mathbf{0}, \tag{13}$$

where $s$ is a learnable scaling factor. This learnable initialization provides an adaptive starting point for the subsequent HQS-based unfolding stages.

### 4.3 HQS-based Sparse Inference Stage

Following the optimization model, the updates of $\mathbf{X}$ and $\mathbf{Z}$ are directly implemented using the analytical solutions in equation 10a and equation 10b. The residual subproblem involves structured error modeling induced by the $\ell_{2,1}$ regularization. Instead of explicitly optimizing its proximal mapping, we parameterize this step with a learnable residual refinement module. Specifically, given the residual representation

$$\mathbf{R}^{t+1} = \mathbf{Y} - \mathbf{D}\mathbf{X}^{t+1}, \tag{14}$$

the residual update is formulated as

$$\mathbf{E}^{t+1} = \mathcal{M}_{\theta_e}(\mathbf{R}^{t+1}) = (1-\rho)\mathbf{R}^{t+1} + \rho\mathcal{N}_{\theta_e}(\mathbf{R}^{t+1}), \tag{15}$$

where $\rho \in [0, 1]$ controls the interpolation between the input residual and the learned refinement, and $\mathcal{N}_{\theta_e}$ denotes the learnable residual refinement network. The network $\mathcal{N}_{\theta_e}$ consists of three $3\times 3$ convolutional layers, where the first two layers are followed by ReLU activation:

$$\mathcal{N}_{\theta_e}(\mathbf{R}) = \mathrm{SNConv}_{3\times3}\left(RELU\left(\mathrm{SNConv}_{3\times3}\left(RELU(\mathrm{SNConv}_{3\times3}(\mathbf{R}))\right)\right)\right), \tag{16}$$

where $\mathrm{SNConv}_{3\times3}$ denotes a spectral-normalized convolutional layer and $RELU(\cdot)$ represents the ReLU activation. The residual interpolation design introduces adaptive refinement while preserving the alternating optimization structure of HQS.

### 4.4 Confidence-aware Prediction Feedback

The proposed closed-loop framework feeds intermediate class predictions back into the subsequent sparse inference process. After each unfolding stage, the reconstructed representation is fed into a shared classifier $f_{\theta_c}(\cdot)$ to obtain the class probability distribution $\mathbf{p}^t$. The class probability is then expanded to the atom level according to the class assignment of dictionary atoms, yielding

Table 1: Dataset characteristics and per-class training budgets. GSD denotes ground sampling distance. Band counts describe the data used before dimensionality reduction.

| | HC | HH | UH18 | SA | TDW | PU | UH13 | QY |
|---|---|---|---|---|---|---|---|---|
| Image size | $1217 \times 303$ | $940 \times 475$ | $601 \times 2385$ | $512 \times 217$ | $1740 \times 860$ | $610 \times 340$ | $349 \times 1905$ | $880 \times 1360$ |
| Bands | 274 | 270 | 48 | 204 | 176 | 103 | 144 | 270 |
| Classes | 16 | 22 | 20 | 16 | 18 | 9 | 15 | 6 |
| GSD (m) | 0.109 | 0.043 | 1.0 | 3.7 | 0.15 | 1.3 | 1.5 | 0.15 |
| Train/class | 50 | 50 | 50 | 10 | 10 | 10 | 20 | 10 |

$\mathbf{q}^t \in \mathbb{R}^K$. To avoid injecting unreliable semantic information during early iterations, we introduce confidence-aware delayed feedback. The effective atom-level semantic response is computed as

$$\mathbf{q}_{\text{eff}}^t = \beta_t \left(1 - \frac{-\sum_{n=1}^{N} p_n^t \log(p_n^t)}{\log N}\right) \mathbf{q}^t, \tag{17}$$

where the term in parentheses defines the normalized entropy-based confidence score of the current prediction, and $\beta_t$ controls the feedback strength at different unfolding stages. We initialize $\beta_1 = 0$, yielding $\mathbf{q}_{\text{eff}}^1 = \mathbf{0}$ for the first unfolding stage. Meanwhile, reconstruction-based evidence is extracted from the similarity between the reconstructed center pixel and dictionary atoms. Specifically, the center pixel representation is denoted as $\mathbf{x}_{\text{center}}^t \in \mathbb{R}^C$, and its similarity with the $K$ dictionary atoms is computed by cosine similarity:

$$S_k^t = \frac{(\mathbf{x}_{\text{center}}^t)^\top \mathbf{d}_k}{\|\mathbf{x}_{\text{center}}^t\|_2 \|\mathbf{d}_k\|_2}, \quad k = 1, \ldots, K, \tag{18}$$

where $\mathbf{d}_k$ denotes the $k$-th dictionary atom. The obtained similarity vector is denoted as $\mathbf{S}^t \in \mathbb{R}^K$. The semantic response $\mathbf{q}_{\text{eff}}^t$ and reconstruction similarity $\mathbf{S}^t$ are concatenated to form an atom-level guidance representation, and the atom-wise relevance weights are generated through a lightweight relevance prediction network:

$$\boldsymbol{\alpha}^t = \mathbf{sig}(\mathbf{W}_2\, ReLU(\mathbf{W}_1([\mathbf{q}_{\text{eff}}^t, \mathbf{S}^t]) + \mathbf{b}_1) + \mathbf{b}_2), \tag{19}$$

where $\mathbf{W}_1, \mathbf{W}_2$ and $\mathbf{b}_1, \mathbf{b}_2$ are learnable parameters, and $\mathbf{sig}(\cdot)$ denotes the sigmoid activation. By combining prediction confidence with reconstruction-aware similarity, the proposed feedback mechanism enables semantic information to progressively participate in sparse inference while retaining reconstruction evidence throughout the unfolding process.

# 5 Experiments

## 5.1 Experimental Setup

**Datasets.** We evaluate our method on eight HSIC benchmarks: Pavia University (PU), University of Houston 2013 (UH13), University of Houston 2018 (UH18), Salinas (SA), WHU-Hi-HanChuan (HC), WHU-Hi-HongHu (HH), QUH-Tangdaowan (TDW), and QUH-Qingyun (QY). These datasets cover diverse vegetation and urban scenes with varying spatial resolutions and spectral characteristics, providing diverse scenarios for evaluating hyperspectral classification performance. Table 1 summarizes their characteristics and training settings. Following common protocols, we randomly select 10 labeled pixels per class for PU, SA, TDW, and QY; 20 for UH2013; and 50 for UH2018, HC, and HH, with all remaining labeled pixels used for testing. We further examine the effectiveness of prediction feedback on ImS using CIFAR-10 and CIFAR-100. For these datasets, we compare open-loop and closed-loop variants under the same unfolding architecture to isolate the effect of the proposed feedback mechanism.

**Comparison Methods.** We compare our method with representative HSIC methods from different architectural families. Specifically, A2S2K (Roy et al., 2021) represents CNN-based approaches; HiT (Yang et al., 2022), SSFTT (Sun et al., 2022), SPRLT (Xue et al., 2022), DBCT (Xu et al., 2024), SQS (Chen et al., 2024), CSCA (Zhang et al., 2025), and SMFAT (Shu et al., 2026) represent Transformer-based approaches; while MambaLG (Pan et al., 2025) represents Mamba-based approaches. These baselines enable a comprehensive evaluation against existing discriminative representation learning paradigms.

Table 2: A quantitative comparison of average OA and Kappa×100 among different HSIC methods, evaluated on eight widely used datasets covering two scenarios: Vegetation (HH, HC, SA, TDW) and Urban (UH18, PU, UH13, QY). Throughout this paper, the best and second-best results of each test case are highlighted in **Bold**, underline, respectively.

| Method | Vegetation | | | | Urban | | | | Avg |
|---|---|---|---|---|---|---|---|---|---|
| | HC | HH | SA | TDW | UH18 | PU | UH13 | QY | |
| A2S2K (Roy et al., 2021) | 87.81<br>85.79 | 88.33<br>85.58 | 91.44<br>90.48 | 73.02<br>70.34 | 72.81<br>66.70 | 89.25<br>86.12 | 90.34<br>89.55 | 73.68<br>66.03 | 83.33<br>80.07 |
| HiT (Yang et al., 2022) | 86.22<br>83.98 | 87.58<br>84.53 | 82.85<br>81.08 | 66.94<br>63.90 | 65.66<br>58.74 | 76.16<br>68.95 | 77.35<br>75.52 | 70.07<br>61.52 | 76.60<br>72.28 |
| SSFTT (Sun et al., 2022) | 87.23<br>85.14 | 80.54<br>76.51 | 93.38<br>92.64 | 72.86<br>70.12 | 70.64<br>64.06 | 85.51<br>81.03 | 90.53<br>89.76 | 66.03<br>56.49 | 80.84<br>76.97 |
| SPRLT (Xue et al., 2022) | 91.49<br>90.07 | 92.35<br>90.43 | 84.13<br>82.29 | 70.37<br>67.07 | 70.01<br>63.85 | 86.30<br>82.16 | 87.51<br>86.52 | 68.04<br>59.50 | 81.27<br>77.74 |
| DBCT (Xu et al., 2024) | 91.88<br>90.55 | 89.73<br>87.29 | 89.02<br>87.78 | 77.40<br>74.90 | 70.18<br>63.68 | 88.39<br>85.19 | 90.61<br>89.85 | 69.03<br>60.49 | 83.28<br>79.96 |
| SQS (Chen et al., 2024) | 87.48<br>85.43 | 86.28<br>82.96 | 93.19<br>92.42 | 71.05<br>67.99 | 73.41<br>66.88 | 87.48<br>83.78 | 86.85<br>85.80 | 68.95<br>60.13 | 81.84<br>78.17 |
| MambaLG (Pan et al., 2025) | 94.23<br>93.27 | 93.51<br>91.85 | 93.47<br>92.75 | 73.67<br>70.96 | 67.99<br>61.59 | 86.04<br>82.23 | 89.85<br>89.02 | 74.00<br>66.61 | 84.10<br>81.03 |
| CSCA (Zhang et al., 2025) | 92.41<br>91.16 | 91.08<br>88.88 | 87.68<br>86.29 | 73.35<br>70.44 | 73.20<br>67.16 | 63.12<br>54.15 | 88.87<br>87.97 | 56.67<br>44.16 | 78.30<br>73.77 |
| SMFAT (Shu et al., 2026) | 93.71<br>92.68 | 95.68<br>94.57 | 95.25<br>94.71 | 76.46<br>73.99 | 75.02<br>69.28 | 89.34<br>86.43 | 89.44<br>88.58 | 73.45<br>66.08 | 86.04<br>83.29 |
| **Pre2Inf (Ours)** | **96.67**<br>**96.11** | **96.37**<br>**95.44** | **97.88**<br>**97.65** | **82.34**<br>**80.30** | **80.31**<br>**74.93** | **91.88**<br>**89.43** | **91.74**<br>**91.06** | **75.72**<br>**68.98** | **89.11**<br>**86.74** |

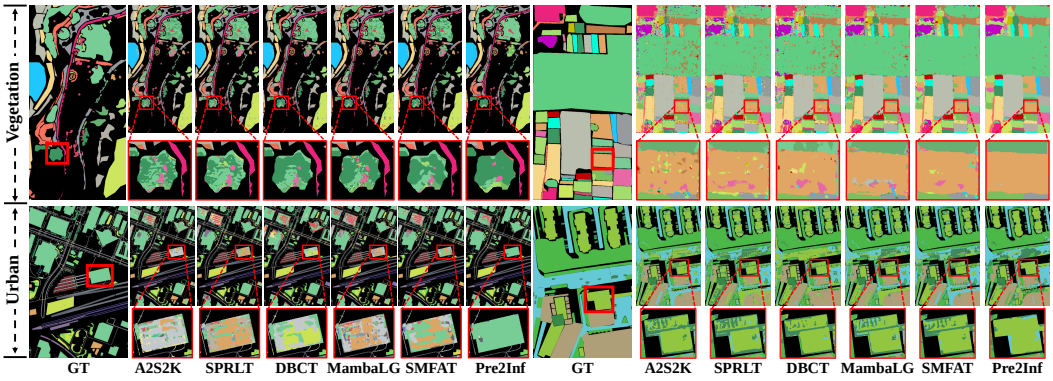



Figure 3: Visualization of classification maps on representative hyperspectral scenes.

**Implementation Details.** To ensure a fair comparison, all competing methods are implemented using their officially reported configurations or publicly released implementations. A fixed random seed is adopted to ensure consistent training and testing splits across different methods. All experiments are conducted with PyTorch on a Linux server equipped with a single NVIDIA H100 GPU. Our method is trained using the Adam optimizer with a batch size of 32, an initial learning rate of 0.001, and 300 epochs. Dataset-specific hyperparameters and loss function settings are provided in the Appendix B to ensure reproducibility. The HSIC performance is evaluated using overall accuracy (OA) and Cohen's kappa coefficient (Kappa). For ImS, accuracy is reported to compare open-loop and closed-loop variants under identical training settings, enabling a controlled evaluation of prediction feedback.

### 5.2 Main Results on Pixel-level Classification

Table 2 presents the comparison between Pre2Inf and representative HSIC methods. Pre2Inf achieves the highest average OA and Kappa of 89.11% and 86.74%, respectively, and obtains the best OA on all eight benchmarks. Compared with the strongest baselines, it improves OA by 2.44%,

Table 3: Stage-wise comparison between open-loop and closed-loop inference. The values in parentheses denote the OA improvements of the closed-loop variant over its open-loop counterpart at the corresponding unfolding stage.

| Dataset | Stage 1 | Stage 2 | Stage 3 | Stage 4 |
|---|---|---|---|---|
| Dataset 8 (Open) | 89.73 | 89.99 | 89.92 | 89.91 |
| Dataset 8 (Closed) | 90.82(**+1.09**) | 91.84(**+1.85**) | 91.84(**+1.92**) | 91.88(**+1.97**) |
| Dataset 9 (Open) | 89.19 | 88.96 | 88.94 | 88.93 |
| Dataset 9 (Closed) | 91.80(**+2.61**) | 91.76(**+2.80**) | 91.74(**+2.79**) | 91.74(**+2.81**) |
| Dataset 10 (Open) | 96.53 | 96.48 | 96.45 | N/A |
| Dataset 10 (Closed) | 97.59(**+1.06**) | 97.55(**+1.07**) | 97.55(**+1.10**) | N/A |
| Dataset 14 (Open) | 80.53 | 80.43 | 80.43 | 80.43 |
| Dataset 14 (Closed) | 82.51(**+1.98**) | 82.36(**+1.93**) | 82.35(**+1.92**) | 82.34(**+1.91**) |

0.69%, 2.63%, and 4.94% on HC, HH, SA, and TDW, respectively, while maintaining consistent gains on urban scenarios. These results demonstrate that prediction feedback effectively enhances sparse inference by enabling intermediate class predictions to regulate subsequent coefficient updates. To further investigate the effectiveness of Pre2Inf, we visualize the classification maps on four representative hyperspectral scenes, including two vegetation scenarios and two urban scenarios. As shown in Figure 3, Pre2Inf produces cleaner and more spatially consistent classification maps compared with existing approaches. In vegetation scenes, where different categories often exhibit complex spectral variations, Pre2Inf preserves intra-class consistency while reducing fragmented predictions. In urban scenes, Pre2Inf better maintains object boundaries and suppresses misclassification noise around buildings and complex structures. These improvements indicate that the proposed prediction-to-inference feedback mechanism enables the unfolding process to exploit evolving class evidence and generate more discriminative representations.

## 5.3 Analysis of Prediction Feedback

### 5.3.1 Closed-Loop Inference

Table 3 compares the open-loop and closed-loop variants across different unfolding stages. The closed-loop design consistently improves OA over the open-loop counterpart, with gains ranging from 1.06% to 2.81% across four datasets. These results demonstrate that feeding intermediate predictions back into subsequent inference can effectively enhance the discriminative capability of sparse optimization. To understand how prediction feedback influences the inference process, we visualize the evolution of prediction confidence and class relevance gap $\Delta\alpha$ in Figure 4. As intermediate predictions become more reliable,the resulting feedback enlarges the relevance distinction between class-related and class-irrelevant atoms, reflected by $\Delta\alpha$, providing adaptive guidance for subsequent coefficient updates. These observations verify that closed-loop unfolding enables predictions to actively regulate the iterative inference trajectory.

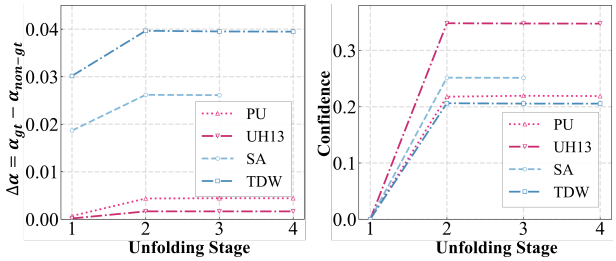



Figure 4: Evolution of prediction feedback signals during closed-loop inference.

### 5.3.2 Prediction-Guided Inference Dynamics

To investigate how prediction feedback should be incorporated into the unfolding process, we compare four variants: Open, Naive, Delay, and Full. As shown in Table 4, Full consistently achieves the best performance across four datasets, improving OA over Open by 1.97%, 2.81%, 1.10%, and 1.91%, respectively. These results demonstrate that properly incorporating intermediate predictions can effectively enhance unfolding inference. The comparisons with Naive and Delay indicate that the feedback strategy influences the inference trajectory. Although Delay improves performance by

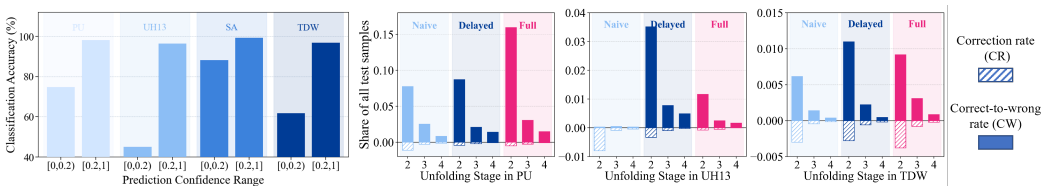



Figure 5: Visualization of prediction feedback dynamics metrics during closed-loop inference.

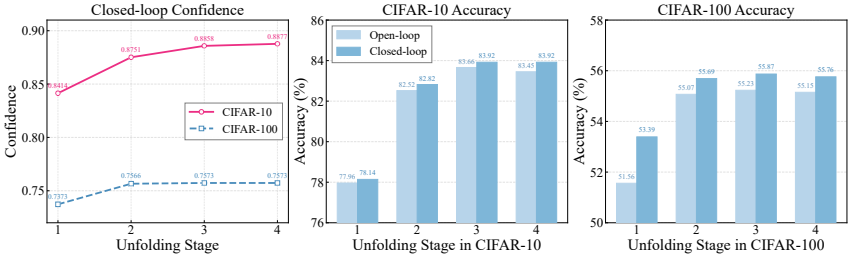



Figure 6: Validation of prediction feedback on ImS using CIFAR-10 and CIFAR-100.

avoiding early unreliable feedback, it cannot achieve the consistent gains of Full across different datasets. Figure 5 further illustrates the inference dynamics of Full. As unfolding proceeds, the correction rate increases while the correct-to-wrong rate remains controlled, indicating that prediction feedback progressively improves coefficient transitions.

### 5.3.3 Cross-task Validation on Image-level Classification

To further verify the generality of prediction feedback, we evaluate the proposed framework on ImS tasks using CIFAR-10 and CIFAR-100. As shown in Table 5, closed-loop inference consistently outperforms open-loop inference, achieving gains of 0.47% and 0.61% on the two datasets, respectively. Figure 6 further shows that prediction confidence progressively increases during unfolding, accompanied by consistent accuracy improvements of the closed-loop variant across stages. These results demonstrate that prediction feedback can effectively improve iterative inference beyond HSIC.

Table 4: Ablation study of different prediction feedback strategies. Open (no prediction feedback), Naive (directly injects predictions), Delay (previous-stage predictions), and Full (Pre2inf).

| Dataset | Open | Naive | Delay | Full |
|---|---|---|---|---|
| 8 | 89.91 | 91.44 | 91.31 | **91.88** |
| 9 | 88.93 | 30.62 | 90.21 | **91.74** |
| 10 | 96.45 | 95.09 | 96.99 | **97.55** |
| 14 | 80.43 | 77.82 | 79.52 | **82.34** |

## 6 Conclusion

We proposed **Pre2Inf**, a closed-loop deep unfolding framework that enables intermediate predictions to guide sparse inference. Unlike conventional open-loop unfolding architectures, Pre2Inf treats intermediate predictions as feedback signals rather than passive outputs, enabling feedback-based regulation of coefficient updates through adaptive relevance weighting. Extensive experiments on hyperspectral and image-level classification benchmarks demonstrate the effectiveness and cross-task generality of the proposed framework. Further analyses reveal that prediction feedback reshapes the inference trajectory by enhancing the discriminability of sparse representations and alleviating ambiguity among competing class-specific representations. Future work will investigate more adaptive feedback mechanisms and extend prediction-as-feedback strategies to broader iterative inference problems.

Table 5: Validation of prediction feedback on ImS.

| Dataset | Open | Closed | Gain |
|---|---|---|---|
| CIFAR-10 | 83.45 | 83.92 | +0.47 |
| CIFAR-100 | 55.15 | 55.76 | +0.61 |

### AI USE STATEMENT

In this work, generative AI tools were used only for language refinement, proofreading, and improving manuscript organization and presentation. We also used these tools to assist with figure layout suggestions and visualization planning.

Generative AI tools were not used to generate experimental results, perform data analysis, develop the core methodology, or formulate scientific claims. All theoretical derivations, algorithm designs, implementations, experiments, and interpretations were conducted and verified by the authors. All AI-assisted suggestions were carefully reviewed and revised by the authors. The authors take full responsibility for the final content of this work.

### ETHICS STATEMENT

This work does not involve human subjects, personal data collection, or any privacy-sensitive information. All datasets used in our experiments are publicly available benchmark datasets that are widely used in the hyperspectral image classification and visual recognition communities. We use these datasets solely for academic research purposes and follow their corresponding usage guidelines. The proposed method focuses on improving model-based iterative inference and does not introduce additional risks related to privacy, security, discrimination, or harmful applications. All experimental results are reported based on the described protocols, and the authors take responsibility for the integrity, reproducibility, and accuracy of the presented research.

### REPRODUCIBILITY STATEMENT

We have made efforts to ensure the reproducibility of this work through detailed descriptions of the proposed framework, experimental protocols, and theoretical analyses. The architecture of Pre2Inf, the closed-loop unfolding procedure, and the optimization formulation are described in the main paper, while additional implementation details, hyperparameter settings, and dataset preprocessing procedures are provided in the Appendix. The stability analysis, including the assumptions and proofs of the theoretical results, is also included in the Appendix. The source code will be released immediately upon acceptance of the paper, while additional experimental details, including hyperparameter settings and implementation configurations, are provided in the Appendix.

## REFERENCES

Rawwad Alhejaili, Motaz Alfarraj, and Ali Al-Shaikhi. Listen to your gradients: Integrating gradients into deep unfolding networks. *Pattern Recognition Letters*, 192:99–105, 2025. ISSN 0167-8655.

Joao Carreira, Pulkit Agrawal, Katerina Fragkiadaki, and Jitendra Malik. Human pose estimation with iterative error feedback. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4733–4742, 2016.

Ning Chen, Leyuan Fang, Yang Xia, Shaobo Xia, Hui Liu, and Jun Yue. Spectral query spatial: Revisiting the role of center pixel in transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024.

Yi Chen, Nasser M. Nasrabadi, and Trac D. Tran. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3973–3985, 2011.

Xili Dai, Ke Chen, Shengbang Tong, Jingyuan Zhang, Xingjian Gao, Mingyang Li, Druv Pai, Yuexiang Zhai, Xiaojun Yuan, Heung-Yeung Shum, Lionel M. Ni, and Yi Ma. Closed-loop transcription via convolutional sparse coding, 2023. URL https://arxiv.org/abs/2302.09347.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*, 2021.

Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the International Conference on Machine Learning*, pp. 399–406, 2010.

Chunming He, Rihan Zhang, Fengyang Xiao, Chengyu Fang, Longxiang Tang, Yulun Zhang, Linghe Kong, Deng-Ping Fan, Kai Li, and Sina Farsiu. RUN: Reversible unfolding network for concealed object segmentation. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 22853–22864. PMLR, 2025.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

John R. Hershey, Jonathan Le Roux, and Felix Weninger. Deep unfolding: Model-based inspiration of novel deep architectures, 2014. URL https://arxiv.org/abs/1409.2574.

Zhuolin Jiang, Zhe Lin, and Larry S. Davis. Label consistent K-SVD: Learning a discriminative dictionary for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (11):2651–2664, 2013.

Chen Liao, Yan Shen, Dan Li, and Zhongli Wang. Using powerful prior knowledge of diffusion model in deep unfolding networks for image compressive sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

Jialin Liu, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. ALISTA: Analytic weights are as good as learned weights in LISTA. In *International Conference on Learning Representations*, 2019.

Julien Mairal, Francis Bach, and Jean Ponce. Task-driven dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2), 2021. doi: 10.1109/MSP.2020.3016905.

Xiangyu Nie, Zhaohui Xue, Hongjun Su, and Jun Li. Nsr-net: Representation model-inspired interpretable deep unfolding network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–16, 2025.

Zhaojie Pan, Chenyu Li, Antonio Plaza, Jocelyn Chanussot, and Danfeng Hong. Hyperspectral image classification with mamba. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–14, 2025.

Patrick Putzky and Max Welling. Recurrent inference machines for solving inverse problems, 2017. URL https://arxiv.org/abs/1706.04008.

Swalpa Kumar Roy, Suvojit Manna, Tiecheng Song, and Lorenzo Bruzzone. Attention-based adaptive spectral–spatial kernel resnet for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7831–7843, 2021.

Zhenqiu Shu, Kexin Zeng, Yuyang Wang, Songze Tang, Zhengtao Yu, and Liang Xiao. Spectral-guided multiscale feature-aware transformer for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 37(5):2193–2207, 2026.

Le Sun, Guangrui Zhao, Yuhui Zheng, and Zebin Wu. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

Kailun Wu, Yiwen Guo, Ziang Li, and Changshui Zhang. Sparse coding with gated learned ISTA. In *International Conference on Learning Representations*, 2020.

Rui Xu, Xue-Mei Dong, Weijie Li, Jiangtao Peng, Weiwei Sun, and Yi Xu. Dbctnet: Double branch convolution-transformer network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024.

Zhaohui Xue, Qi Xu, and Mengxue Zhang. Local transformer with spatial partition restore for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:4307–4325, 2022.

Xiaofei Yang, Weijia Cao, Yao Lu, and Yicong Zhou. Hyperspectral image transformer classification networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.

Yaodong Yu, Sam Buchanan, Druv Pai, Tianzhe Chu, Ziyang Wu, Shengbang Tong, Hao Bai, Yuexiang Zhai, Benjamin D. Haeffele, and Yi Ma. White-box transformers via sparse rate reduction: Compression is all there is? *Journal of Machine Learning Research*, 25(300):1–128, 2024.

Amir R. Zamir, Te-Lin Wu, Lin Sun, William Shen, Jitendra Malik, and Silvio Savarese. Feedback networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

Bo Zhang, Yaxiong Chen, Shengwu Xiong, and Xiaoqiang Lu. Hyperspectral image classification via cascaded spatial cross-attention network. *IEEE Transactions on Image Processing*, 34:899–913, 2025.

Jian Zhang and Bernard Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

# A Stability Analysis of Closed-Loop Inference

This section analyzes the stability properties of the proposed prediction-feedback unfolding process. We first establish the well-posedness of the coefficient update and then characterize the sensitivity of the update with respect to prediction-induced relevance perturbations. Finally, we provide a conditional convergence result for the complete iterative system.

## A.1 Well-Posedness of Prediction-Conditioned Coefficient Updates

The relevance weights generated by the feedback module satisfy

$$0 \le \alpha_k^t \le 1, \quad k = 1, \dots, K, \tag{20}$$

because the relevance generator employs a sigmoid activation. Therefore, the prediction-conditioned regularization term

$$\frac{\gamma}{2} \|(\mathbf{1} - \boldsymbol{\alpha}^t)^{1/2} \odot \mathbf{X}\|_F^2 \tag{21}$$

is always non-negative.

For fixed $\boldsymbol{\alpha}^t$, the coefficient update is determined by

$$\mathbf{A}^t \mathbf{X}^{t+1} = \mathbf{D}^\top (\mathbf{Y} - \mathbf{E}^t) + \eta \mathbf{Z}^t, \tag{22}$$

where

$$\mathbf{A}^t = \mathbf{D}^\top \mathbf{D} + \eta \mathbf{I} + \gamma \operatorname{Diag}(\mathbf{1} - \boldsymbol{\alpha}^t). \tag{23}$$

Since $\mathbf{D}^\top \mathbf{D} \succeq 0$ and $\operatorname{Diag}(\mathbf{1} - \boldsymbol{\alpha}^t) \succeq 0$,

$$\mathbf{A}^t \succeq \eta \mathbf{I}. \tag{24}$$

Therefore, $\mathbf{A}^t$ is positive definite and the coefficient subproblem admits a unique solution.

**Proposition 1.** For fixed $\boldsymbol{\alpha}^t$, $\eta > 0$, and $\gamma \ge 0$, the coefficient update has a unique solution and satisfies

$$\|(\mathbf{A}^t)^{-1}\|_2 \le \frac{1}{\eta}. \tag{25}$$

**Proof.** For any nonzero vector $\mathbf{v}$,

$$\begin{aligned} \mathbf{v}^\top \mathbf{A}^t \mathbf{v} =& \|\mathbf{D}\mathbf{v}\|_2^2 + \eta \|\mathbf{v}\|_2^2 \\ &+ \gamma \sum_k (1 - \alpha_k^t) v_k^2 \\ \ge& \eta \|\mathbf{v}\|_2^2 > 0. \end{aligned} \tag{26}$$

Hence $\mathbf{A}^t$ is positive definite, which proves uniqueness and the inverse bound. □

### A.2 Stability under prediction-induced relevance perturbations

Prediction feedback influences coefficient estimation through $\boldsymbol{\alpha}^t$. We analyze how perturbations in the feedback signal affect the coefficient update. Let $\mathbf{X}^+$ and $\widetilde{\mathbf{X}}^+$ denote coefficient updates obtained from $(\mathbf{Z}, \mathbf{E}, \boldsymbol{\alpha})$ and $(\widetilde{\mathbf{Z}}, \widetilde{\mathbf{E}}, \widetilde{\boldsymbol{\alpha}})$, respectively.

**Proposition 2.** The difference between the two coefficient updates satisfies

$$\begin{aligned}\|\mathbf{X}^+ - \widetilde{\mathbf{X}}^+\|_F \leq & \|\mathbf{Z} - \widetilde{\mathbf{Z}}\|_F + \frac{\|\mathbf{D}\|_2}{\eta}\|\mathbf{E} - \widetilde{\mathbf{E}}\|_F \\ & + \frac{\gamma}{\eta}\|\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}\|_\infty \|\widetilde{\mathbf{X}}^+\|_F.\end{aligned} \tag{27}$$

**Proof.** The coefficient updates satisfy

$$\mathbf{A}\mathbf{X}^+ = \mathbf{D}^\top(\mathbf{Y} - \mathbf{E}) + \eta\mathbf{Z}, \tag{28}$$

and

$$\widetilde{\mathbf{A}}\widetilde{\mathbf{X}}^+ = \mathbf{D}^\top(\mathbf{Y} - \widetilde{\mathbf{E}}) + \eta\widetilde{\mathbf{Z}}, \tag{29}$$

where

$$\mathbf{A} = \mathbf{D}^\top\mathbf{D} + \eta\mathbf{I} + \gamma\,\mathrm{Diag}(\mathbf{1} - \boldsymbol{\alpha}), \tag{30}$$

and $\widetilde{\mathbf{A}}$ is defined similarly.

Subtracting the two equations gives

$$\begin{aligned}\mathbf{A}(\mathbf{X}^+ - \widetilde{\mathbf{X}}^+) = & \eta(\mathbf{Z} - \widetilde{\mathbf{Z}}) - \mathbf{D}^\top(\mathbf{E} - \widetilde{\mathbf{E}}) \\ & + (\widetilde{\mathbf{A}} - \mathbf{A})\widetilde{\mathbf{X}}^+.\end{aligned} \tag{31}$$

Since

$$\widetilde{\mathbf{A}} - \mathbf{A} = \gamma\,\mathrm{Diag}(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}), \tag{32}$$

multiplying both sides by $\mathbf{A}^{-1}$ and using Proposition 1 yields

$$\begin{aligned}\|\mathbf{X}^+ - \widetilde{\mathbf{X}}^+\|_F \leq & \frac{1}{\eta}\|\eta(\mathbf{Z} - \widetilde{\mathbf{Z}}) - \mathbf{D}^\top(\mathbf{E} - \widetilde{\mathbf{E}})\|_F \\ & + \frac{\gamma}{\eta}\|\,\mathrm{Diag}(\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}})\widetilde{\mathbf{X}}^+\|_F.\end{aligned} \tag{33}$$

Applying the sub-multiplicative property of matrix norms gives

$$\begin{aligned}\|\mathbf{X}^+ - \widetilde{\mathbf{X}}^+\|_F \leq & \|\mathbf{Z} - \widetilde{\mathbf{Z}}\|_F + \frac{\|\mathbf{D}\|_2}{\eta}\|\mathbf{E} - \widetilde{\mathbf{E}}\|_F \\ & + \frac{\gamma}{\eta}\|\boldsymbol{\alpha} - \widetilde{\boldsymbol{\alpha}}\|_\infty \|\widetilde{\mathbf{X}}^+\|_F,\end{aligned} \tag{34}$$

which completes the proof. □

### A.3 Conditional stability of closed-loop inference dynamics

The proposed network uses a finite number of unfolding stages and does not require convergence during inference. Nevertheless, We further characterize the behavior of an indefinitely repeated stationary version of the proposed inference process.

Let

$$\mathbf{U}^t = (\mathbf{X}^t, \mathbf{Z}^t, \mathbf{E}^t) \tag{35}$$

denote the complete optimization state, and let

$$\mathbf{U}^{t+1} = \mathcal{T}(\mathbf{U}^t) \tag{36}$$

represent the complete feedback update. **Theorem 1.** Assume that a closed set $\Omega$ satisfies $\mathcal{T}(\Omega) \subseteq \Omega$ and that

$$\|\mathcal{T}(\mathbf{U}) - \mathcal{T}(\mathbf{V})\|_F \leq q\|\mathbf{U} - \mathbf{V}\|_F, \quad 0 \leq q < 1. \tag{37}$$

Then the repeated prediction-feedback updates converge to a unique fixed point.

**Proof.** The state space is finite-dimensional and $\Omega$ is complete. Therefore, the contraction mapping theorem guarantees the existence and uniqueness of a fixed point and geometric convergence:

$$\|\mathbf{U}^t - \mathbf{U}^*\|_F \leq q^{t-1}\|\mathbf{U}^1 - \mathbf{U}^*\|_F. \tag{38}$$

□

This result provides a conditional characterization of the closed-loop inference dynamics. The finite-stage network used in experiments does not rely on this convergence assumption.

# B EXPERIMENTS

This section provides additional experimental details and analyses to further validate the effectiveness and reproducibility of Pre2Inf. We first describe the training objective. Then, we conduct comprehensive parameter analyses to investigate the sensitivity of the proposed framework to different configurations. Ablation studies are further provided to analyze the contribution of each component. Finally, we report detailed comparisons with representative hyperspectral image classification methods on all benchmark datasets.

## B.1 TRAINING OBJECTIVE

The proposed network is trained by jointly optimizing classification supervision, HQS consistency, and class-aware reconstruction. The overall objective is defined as

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_h \mathcal{L}_{hqs} + \lambda_r \mathcal{L}_{rec}, \tag{39}$$

where $\mathcal{L}_{cls}$, $\mathcal{L}_{hqs}$, and $\mathcal{L}_{rec}$ denote the classification, unfolding consistency, and class-aware reconstruction losses, respectively. The classification loss supervises intermediate predictions from all unfolding stages:

$$\mathcal{L}_{cls} = \sum_{t=1}^{T} w_t \left[ -\sum_{n=1}^{N} (1 - p_n^t)^{\gamma} y_n \log(p_n^t) \right], \tag{40}$$

where $y_n$ is the one-hot label, $p_n^t$ is the predicted probability of class $n$ at stage $t$, $\gamma$ is the focal loss parameter, and $w_t$ is the stage-wise weight. To preserve the sparse optimization structure during unfolding, we introduce the HQS consistency loss:

$$\mathcal{L}_{hqs} = \sum_{t=1}^{T} w_t \left( \lambda_x \|\mathbf{X}^t\|_1 + \lambda_e \|\mathbf{E}^t\|_{2,1} + \|\mathbf{Y} - \mathbf{D}\mathbf{X}^t - \mathbf{E}^t\|_F^2 \right). \tag{41}$$

Furthermore, considering the class-structured dictionary, we introduce a class-aware reconstruction loss to encourage dictionary atoms associated with the ground-truth class to achieve more faithful reconstruction:

$$\mathcal{L}_{rec} = -\sum_{c=1}^{N} y_c \log \frac{\exp\left(-\frac{e_c}{\tau_r \bar{e}}\right)}{\sum_{j=1}^{N} \exp\left(-\frac{e_j}{\tau_r \bar{e}}\right)}, \tag{42}$$

where $e_c$ denotes the reconstruction error of class $c$, $\bar{e}$ is the sample-wise normalization factor, $\tau_r$ is the temperature parameter, and $y_c$ is the class indicator. This loss encourages the ground-truth class dictionary to provide lower reconstruction error than competing classes.

## B.2 HYPERPARAMETER SENSITIVITY ANALYSIS

We evaluate the sensitivity of the proposed framework to three key hyperparameters, including patch size, the number of unfolding stages, and the number of atoms per class. As shown in Fig. **??**, the proposed method exhibits stable performance across different configurations. Based on the overall accuracy trends, we select a patch size of 6, 4 unfolding stages, and 3 atoms per class as the default settings. These configurations provide a good trade-off between representation capacity and inference efficiency.

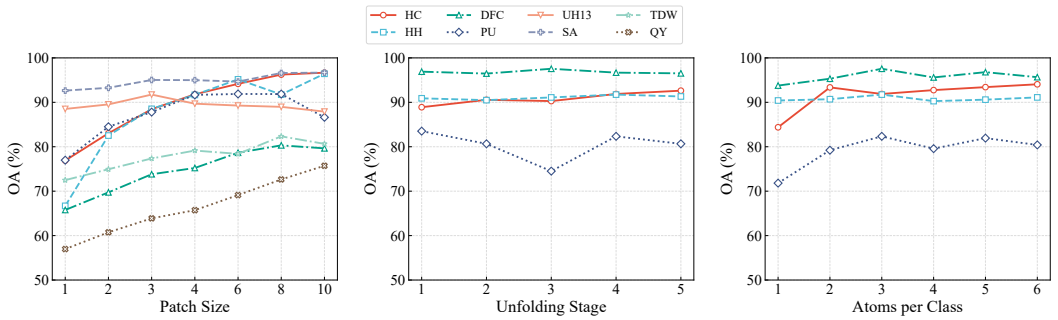



Figure 7: Sensitivity analysis of key hyperparameters, including patch size, unfolding stages, and atoms per class. The proposed framework maintains stable classification performance across different parameter settings.

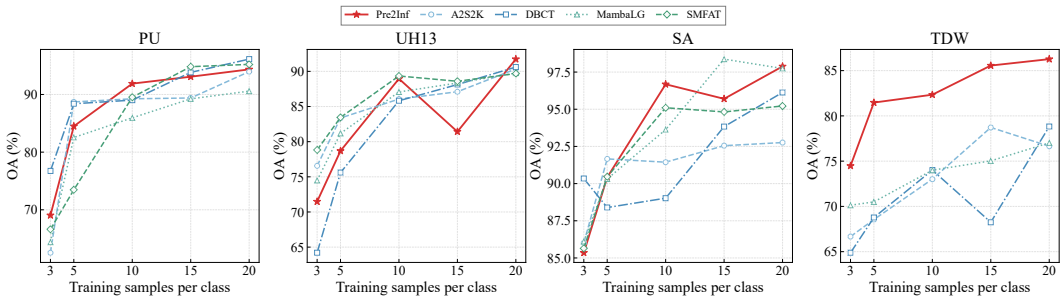



Figure 8: Influence of the number of training samples on classification performance. The results are evaluated on PU, UH13, SA, and TDW datasets with different numbers of labeled samples per class.

We further investigate the influence of the number of labeled training samples. Fig. 8 compares different methods under varying sample budgets. The proposed method maintains competitive performance with limited training samples and consistently improves as more labeled samples become available, demonstrating its robustness in data-limited scenarios.

### B.3 Component-wise Ablation Analysis

To further investigate the contribution of individual components in the proposed framework, we conduct component-wise ablation studies on four representative hyperspectral datasets. Starting from the full model, each variant removes one specific component while keeping other modules unchanged. The evaluated components include the dictionary prior ($D$), initialization strategy ($X^0$), error modeling ($E$), reconstruction–atom similarity guidance ($sim$), prediction feedback ($feed$), confidence modulation, and reconstruction constraint ($recon$).

As reported in Table 6, removing any individual component leads to performance degradation on most datasets, demonstrating that each component contributes to the effectiveness of the proposed framework. In particular, removing the prediction feedback module results in consistent accuracy drops, highlighting the importance of the closed-loop inference mechanism. The confidence module further improves robustness by regulating the influence of intermediate predictions, while similarity guidance and reconstruction constraints provide complementary structural information for discriminative representation learning. The complete model achieves the best performance across all evaluated datasets, indicating that the proposed components jointly establish an effective prediction-guided unfolding framework.

### B.4 Per-class Classification Results

We provide detailed classification results on eight hyperspectral benchmarks, including OA, AA, Kappa, and per-class accuracies. These results offer a comprehensive evaluation of the proposed

Table 6: Component-wise ablation study on four hyperspectral datasets. Each variant removes one component from the full model while keeping other modules unchanged.

| Dataset | $D$ | $X^0$ | $E$ | $sim$ | $feed$ | Confidence | Recon | Full |
|---|---|---|---|---|---|---|---|---|
| PU | 90.66 | 90.30 | 88.09 | 91.05 | 89.91 | 91.31 | 90.79 | **91.88** |
| UH13 | 90.69 | 88.11 | 90.26 | 91.04 | 88.93 | 90.21 | 91.16 | **91.74** |
| SA | 89.47 | 97.25 | 97.33 | 87.40 | 96.87 | 86.52 | 97.35 | **97.55** |
| TDW | 80.07 | 79.25 | 79.88 | 80.48 | 80.43 | 79.52 | 79.89 | **82.34** |

Table 7: Detailed classification results on the HC dataset. Best and second-best results are highlighted in bold and underline, respectively.

| Metric | A2S2K | CSCA | DBCT | HiT | MambaLG | SPRLT-Net | SQSFormer | SMFAT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| OA(%) | 87.81 | 92.41 | 91.88 | 86.22 | <u>94.23</u> | 91.49 | 87.48 | 93.71 | **96.67** |
| AA(%) | 78.99 | 86.11 | 83.72 | 74.12 | **94.29** | 83.13 | 76.47 | 90.21 | <u>93.68</u> |
| Kappa(%) | 85.79 | 91.16 | 90.55 | 83.98 | <u>93.27</u> | 90.07 | 85.43 | 92.68 | **96.11** |
| Class1 | 93.30 | 97.18 | <u>98.70</u> | 95.31 | 95.12 | 97.78 | 95.77 | 98.39 | **99.02** |
| Class2 | 95.22 | <u>97.91</u> | 96.56 | 88.91 | 84.99 | 88.46 | 93.89 | 96.24 | **98.09** |
| Class3 | 69.40 | 98.03 | 85.73 | 79.05 | **98.64** | 89.12 | 76.23 | 94.36 | <u>98.56</u> |
| Class4 | 94.54 | 98.65 | 95.24 | 87.60 | 97.45 | 99.01 | 96.50 | <u>99.28</u> | **99.98** |
| Class5 | 53.05 | 76.08 | 66.61 | 38.30 | **100.00** | 59.96 | 51.05 | 87.58 | <u>92.51</u> |
| Class6 | 44.84 | 78.55 | 67.99 | 41.18 | **93.17** | 62.42 | 56.50 | 75.49 | <u>82.16</u> |
| Class7 | 75.39 | 80.82 | 81.59 | 71.68 | <u>96.16</u> | 81.29 | 67.65 | 90.14 | **97.66** |
| Class8 | 94.59 | 88.23 | 93.66 | 81.54 | 89.35 | 84.31 | 84.22 | **97.81** | <u>97.13</u> |
| Class9 | 72.49 | 84.95 | 79.72 | 73.36 | **91.21** | 75.67 | 67.09 | 68.57 | <u>87.95</u> |
| Class10 | 91.13 | 96.32 | 96.93 | 94.89 | **98.38** | 96.94 | 96.49 | 96.80 | <u>97.89</u> |
| Class11 | 87.99 | 80.45 | 89.59 | 80.94 | **94.30** | 92.66 | 85.77 | 90.57 | <u>92.86</u> |
| Class12 | 73.36 | 81.01 | 74.12 | 55.08 | **98.98** | 77.11 | 53.39 | <u>97.61</u> | 96.98 |
| Class13 | 62.35 | 61.69 | 65.16 | 55.46 | **84.47** | 66.13 | 67.93 | 64.02 | <u>79.96</u> |
| Class14 | 84.08 | **97.55** | 85.34 | 84.89 | 92.25 | 93.81 | 77.03 | 95.84 | <u>96.44</u> |
| Class15 | 73.18 | 60.37 | 62.62 | 58.09 | **96.31** | 65.71 | 54.28 | <u>90.75</u> | 81.93 |
| Class16 | 98.91 | 99.94 | **99.99** | 99.67 | 97.84 | 99.75 | 99.73 | <u>99.98</u> | 99.81 |

method across different datasets and semantic categories. Table 7–Table 14 summarize the detailed classification results on HH, HC, UH18, PU, UH13, SA, TDW, and QY, respectively.

Table 8: Detailed classification results on the HH dataset. Best and second-best results are highlighted in bold and underline, respectively.

| Metric | A2S2K | CSCA | DBCT | HiT | MambaLG | SPRLT-Net | SQSFormer | SMFAT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| OA(%) | 88.33 | 91.08 | 89.73 | 87.58 | 93.51 | 92.35 | 86.28 | 95.68 | **96.37** |
| AA(%) | 78.37 | 84.18 | 84.29 | 74.58 | **94.06** | 82.98 | 73.31 | 90.18 | 91.74 |
| Kappa(%) | 85.58 | 88.88 | 87.29 | 84.53 | 91.85 | 90.43 | 82.96 | 94.57 | **95.44** |
| Class1 | 96.68 | **99.06** | 97.21 | 98.34 | 96.35 | 97.41 | 95.40 | 97.17 | 95.22 |
| Class2 | 83.49 | 76.08 | 88.83 | 75.45 | **94.68** | 81.93 | 72.67 | 81.66 | 71.07 |
| Class3 | 90.30 | 96.74 | 87.13 | 96.50 | 92.73 | 98.64 | 96.44 | **98.74** | 98.46 |
| Class4 | 99.42 | 99.25 | 99.65 | 99.41 | 96.77 | 99.63 | 99.22 | **99.79** | 99.70 |
| Class5 | 28.71 | 36.88 | 31.48 | 38.84 | **97.45** | 48.59 | 33.37 | 67.66 | 66.15 |
| Class6 | 97.78 | 98.83 | 98.62 | 96.84 | 91.50 | 99.37 | 98.09 | 99.56 | **99.82** |
| Class7 | 95.10 | 96.14 | 92.76 | 94.86 | 75.22 | 96.92 | 91.34 | **98.60** | 97.78 |
| Class8 | 62.33 | 51.45 | 78.51 | 30.11 | **93.28** | 45.76 | 40.88 | 58.32 | 82.17 |
| Class9 | 99.01 | 97.54 | **99.67** | 99.09 | 98.18 | 98.70 | 99.31 | 99.28 | 98.36 |
| Class10 | 86.13 | 95.36 | 96.71 | 79.83 | 88.67 | 86.79 | 94.26 | **98.74** | 98.70 |
| Class11 | 87.41 | 91.06 | 96.29 | 72.12 | 92.51 | 86.59 | 79.68 | 91.21 | **97.43** |
| Class12 | 64.33 | 63.03 | 65.44 | 62.86 | **93.29** | 74.25 | 43.93 | 87.19 | 92.75 |
| Class13 | 83.78 | 79.96 | 79.75 | 76.35 | 87.82 | **88.40** | 75.38 | 87.31 | 87.97 |
| Class14 | 77.64 | 91.10 | 95.92 | 86.09 | 94.50 | 89.08 | 86.56 | **97.68** | 95.07 |
| Class15 | 38.01 | 78.93 | 93.63 | 62.59 | **98.95** | 82.92 | 48.68 | 89.38 | 82.48 |
| Class16 | 95.24 | 98.72 | 78.94 | 98.36 | 99.39 | 97.89 | 98.56 | 99.12 | **99.87** |
| Class17 | 89.11 | 91.81 | **99.92** | 74.06 | 91.72 | 94.80 | 80.93 | 99.34 | 97.24 |
| Class18 | 80.22 | 86.85 | 77.07 | 60.64 | **96.56** | 69.16 | 42.45 | 87.92 | 86.92 |
| Class19 | 83.74 | 89.71 | 82.95 | 84.24 | 92.43 | 89.59 | 82.96 | **93.58** | 92.67 |
| Class20 | 74.83 | 88.86 | 82.08 | 67.46 | **98.54** | 92.99 | 57.10 | 95.52 | 96.82 |
| Class21 | 52.78 | 72.73 | 76.58 | 25.84 | **99.92** | 39.82 | 34.02 | 82.98 | 88.74 |
| Class22 | 58.22 | 71.84 | 55.34 | 60.81 | **98.85** | 66.28 | 61.51 | 73.24 | 92.85 |

Table 9: Detailed classification results on the UH18 dataset. Best and second-best results are highlighted in bold and underline, respectively.

| Metric | A2S2K | CSCA | DBCT | HiT | MambaLG | SPRLT-Net | SQSFormer | SMFAT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| OA(%) | 72.81 | 73.20 | 70.18 | 65.66 | 67.99 | 70.01 | 73.41 | 75.02 | **80.31** |
| AA(%) | 61.63 | 67.25 | 61.58 | 51.55 | **84.86** | 55.32 | 60.52 | 62.47 | 64.89 |
| Kappa(%) | 66.70 | 67.16 | 63.68 | 58.74 | 61.59 | 63.85 | 66.88 | 69.28 | **74.93** |
| Class1 | 77.11 | 55.88 | 55.65 | 69.40 | **89.52** | 82.45 | 48.35 | 56.78 | 51.25 |
| Class2 | **91.87** | 89.50 | 84.23 | 83.64 | 89.60 | 90.20 | 85.29 | 84.59 | 86.68 |
| Class3 | 76.81 | 99.66 | 78.41 | 49.83 | **100.00** | 57.07 | 89.98 | 82.36 | 83.99 |
| Class4 | 84.88 | 77.46 | 86.32 | 85.94 | **97.63** | 86.58 | 71.73 | 85.20 | 69.89 |
| Class5 | 52.43 | 46.44 | 47.66 | 43.67 | **91.12** | 27.14 | 41.82 | 42.40 | 36.48 |
| Class6 | 54.59 | 76.99 | 73.83 | 26.60 | **96.95** | 35.43 | 62.15 | 47.23 | 56.47 |
| Class7 | 59.74 | **100.00** | 11.82 | 60.71 | **100.00** | 50.00 | 69.78 | 58.44 | 12.59 |
| Class8 | 64.97 | 65.15 | 65.69 | 61.28 | **86.39** | 69.40 | 61.76 | 73.40 | 82.36 |
| Class9 | 96.93 | 97.90 | 95.71 | 95.80 | 63.86 | 97.39 | 95.51 | **98.09** | 96.71 |
| Class10 | 63.46 | 59.87 | 70.49 | 48.97 | 42.28 | 60.96 | 58.68 | 60.38 | **75.20** |
| Class11 | 42.17 | 40.31 | 43.26 | 44.02 | 41.08 | **49.22** | 44.71 | 45.63 | 48.74 |
| Class12 | 6.70 | 8.10 | 6.86 | 6.37 | **72.65** | 5.06 | 8.09 | 8.37 | 8.50 |
| Class13 | 69.13 | 63.34 | 69.41 | 52.71 | 55.32 | 72.08 | **78.48** | 73.39 | 77.87 |
| Class14 | 45.29 | 54.58 | 55.00 | 47.63 | **91.63** | 60.44 | 46.03 | 48.72 | 54.60 |
| Class15 | 65.73 | 97.55 | 87.90 | 47.55 | **97.79** | 42.30 | 84.85 | 78.02 | 93.20 |
| Class16 | 69.87 | 80.70 | 80.95 | 39.48 | **93.01** | 57.57 | 80.97 | 69.11 | 83.62 |
| Class17 | 26.19 | 50.57 | 69.29 | 3.82 | **100.00** | 3.25 | 6.85 | 32.35 | 24.79 |
| Class18 | 37.99 | 36.79 | 20.58 | 46.37 | **94.68** | 35.20 | 42.79 | 48.49 | 81.62 |
| Class19 | 76.76 | 49.94 | 47.29 | 45.72 | **94.53** | 52.87 | 62.29 | 79.20 | 82.40 |
| Class20 | 69.88 | 94.37 | 81.27 | 71.58 | **99.14** | 71.86 | 70.30 | 77.28 | 90.90 |

Table 10: Detailed classification results on the PU dataset. Best and second-best results are highlighted in bold and underline, respectively.

| Metric | A2S2K | CSCA | DBCT | HiT | MambaLG | SPRLT-Net | SQSFormer | SMFAT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| OA(%) | 89.25 | 63.12 | 88.39 | 76.16 | 86.04 | 86.30 | 87.48 | 89.34 | **91.88** |
| AA(%) | 86.45 | 55.37 | 86.06 | 73.15 | **90.74** | 84.19 | 82.46 | 88.72 | 86.86 |
| Kappa(%) | 86.12 | 54.15 | 85.19 | 68.95 | 82.23 | 82.16 | 83.78 | 86.43 | **89.43** |
| Class1 | 96.19 | 53.23 | 87.38 | 87.31 | 79.40 | 91.42 | 95.72 | **98.14** | 95.42 |
| Class2 | 97.30 | 93.32 | 99.52 | 86.77 | 82.73 | 91.87 | 97.44 | 99.22 | **99.58** |
| Class3 | 68.31 | 35.58 | 60.88 | 44.97 | **96.86** | 64.33 | 50.71 | 88.58 | 72.19 |
| Class4 | **97.82** | 58.88 | 77.26 | 97.13 | 88.44 | 96.66 | 93.59 | 92.49 | 95.36 |
| Class5 | 99.93 | 85.02 | 99.78 | 96.53 | 99.78 | 99.56 | **100.00** | 99.20 | 98.26 |
| Class6 | 72.57 | 57.98 | 80.36 | 52.35 | **95.29** | 67.85 | 84.74 | 63.16 | 88.75 |
| Class7 | 64.60 | 40.25 | **95.08** | 37.83 | 94.35 | 68.57 | 80.39 | 73.73 | 72.16 |
| Class8 | 87.15 | 35.59 | 86.52 | 72.02 | 82.67 | 87.34 | 66.37 | **90.66** | 76.41 |
| Class9 | 94.15 | 38.50 | 87.77 | 83.47 | **97.14** | 90.08 | 73.23 | 93.26 | 83.57 |

Table 11: Detailed classification results on the UH13 dataset. Best and second-best results are highlighted in bold and underline, respectively.

| Metric | A2S2K | CSCA | DBCT | HiT | MambaLG | SPRLT-Net | SQSFormer | SMFAT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| OA(%) | 90.34 | 88.87 | 90.61 | 77.35 | 89.85 | 87.51 | 86.85 | 89.44 | **91.74** |
| AA(%) | 92.45 | 90.03 | 91.76 | 82.62 | 91.61 | 89.21 | 87.88 | 91.51 | **92.82** |
| Kappa(%) | 89.55 | 87.97 | 89.85 | 75.52 | 89.02 | 86.52 | 85.80 | 88.58 | **91.06** |
| Class1 | 93.63 | 87.73 | 90.01 | 55.61 | **98.78** | 80.72 | 92.77 | 80.60 | 89.21 |
| Class2 | 99.11 | 97.79 | 92.04 | 95.21 | 96.35 | **99.91** | 92.94 | **99.91** | 98.02 |
| Class3 | **100.00** | 99.41 | **100.00** | 99.68 | **100.00** | 50.64 | 99.85 | 99.12 | 98.69 |
| Class4 | 99.22 | 95.36 | 96.29 | 96.72 | 94.03 | 94.88 | 98.82 | 97.71 | **99.91** |
| Class5 | 98.29 | 99.92 | **100.00** | 97.04 | 99.92 | **100.00** | 99.19 | **100.00** | 99.67 |
| Class6 | 92.83 | 80.42 | 99.00 | 65.85 | **99.34** | 90.48 | 92.83 | 98.06 | 92.55 |
| Class7 | 77.57 | 79.91 | 84.15 | 84.86 | 87.09 | **93.53** | 86.02 | 85.82 | 85.76 |
| Class8 | **100.00** | 97.36 | 88.94 | 93.30 | 71.87 | 91.10 | 99.18 | 86.10 | 97.76 |
| Class9 | 94.39 | 87.47 | 89.44 | 72.19 | 81.15 | 90.46 | 90.48 | **95.67** | 92.11 |
| Class10 | 82.46 | 85.11 | **88.55** | 74.36 | 72.89 | 76.33 | 75.77 | 81.35 | 81.76 |
| Class11 | 71.86 | 72.44 | 77.25 | 88.04 | 93.08 | **97.07** | 72.43 | 72.68 | 77.90 |
| Class12 | 81.99 | 84.74 | **92.25** | 49.56 | 84.49 | 87.34 | 72.73 | 87.69 | 90.03 |
| Class13 | **96.81** | 88.24 | 80.00 | 80.21 | 95.54 | 90.11 | 54.61 | 92.34 | 94.98 |
| Class14 | **100.00** | 97.37 | **100.00** | 96.58 | **100.00** | **100.00** | 99.03 | **100.00** | 94.00 |
| Class15 | 98.61 | 97.11 | 98.46 | 90.13 | 99.69 | 95.63 | 91.55 | 95.62 | **100.00** |

Table 12: Detailed classification results on the SA dataset. Best and second-best results are highlighted in bold and underline, respectively.

| Metric | A2S2K | CSCA | DBCT | HiT | MambaLG | SPRLT-Net | SQSFormer | SMFAT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| OA(%) | 91.44 | 87.68 | 89.02 | 82.85 | 93.47 | 84.13 | 93.19 | 95.25 | **97.88** |
| AA(%) | 96.19 | 85.64 | 94.81 | 91.50 | 96.75 | 92.22 | 94.87 | **97.47** | 97.12 |
| Kappa(%) | 90.48 | 86.29 | 87.78 | 81.08 | 92.75 | 82.29 | 92.42 | 94.71 | **97.65** |
| Class1 | **100.00** | 80.34 | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** | **100.00** |
| Class2 | **100.00** | 93.09 | **100.00** | 97.00 | 99.89 | 99.68 | 99.84 | **100.00** | **100.00** |
| Class3 | 98.93 | 87.17 | **100.00** | 95.89 | 99.39 | 97.40 | 99.59 | 98.89 | **100.00** |
| Class4 | 99.57 | 76.17 | 93.89 | 94.85 | **100.00** | 95.71 | 88.18 | 92.88 | 71.14 |
| Class5 | 99.77 | 90.50 | **100.00** | 99.50 | 97.45 | **100.00** | 96.92 | **100.00** | 99.96 |
| Class6 | **100.00** | 98.20 | **100.00** | 99.77 | 99.87 | **100.00** | 99.97 | **100.00** | 99.90 |
| Class7 | **100.00** | 93.02 | 98.81 | 96.31 | 99.61 | 95.32 | 98.86 | 99.27 | 98.28 |
| Class8 | 82.72 | 87.85 | 74.39 | 74.44 | 82.03 | 62.10 | 86.62 | 90.99 | **99.85** |
| Class9 | 99.58 | 94.91 | 99.25 | 99.71 | **100.00** | 99.60 | 99.97 | 99.61 | 99.65 |
| Class10 | 95.91 | 91.67 | 90.04 | 85.17 | 95.01 | 83.49 | 96.80 | 95.60 | **98.78** |
| Class11 | 99.91 | 87.49 | 99.81 | 88.32 | 98.58 | 98.10 | 97.42 | **100.00** | **100.00** |
| Class12 | **100.00** | 97.29 | **100.00** | 98.20 | **100.00** | **100.00** | 93.07 | **100.00** | 98.99 |
| Class13 | 99.78 | 62.32 | 99.89 | 99.67 | **100.00** | 99.67 | 87.44 | 99.23 | 93.89 |
| Class14 | 99.52 | 61.60 | 97.22 | 94.08 | 99.34 | 99.33 | 97.19 | **99.91** | 98.17 |
| Class15 | 68.52 | 82.55 | 63.94 | 46.99 | 85.49 | 46.76 | 81.16 | 83.45 | **95.39** |
| Class16 | 94.78 | 86.00 | 99.67 | 94.15 | 91.26 | 98.43 | 94.81 | 99.61 | **100.00** |

Table 13: Detailed classification results on the TDW dataset. Best and second-best results are highlighted in bold and underline, respectively.

| Metric | A2S2K | CSCA | DBCT | HiT | MambaLG | SPRLT-Net | SQSFormer | SMFAT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| OA(%) | 73.02 | 73.35 | 77.40 | 66.94 | 73.67 | 70.37 | 71.05 | 76.46 | **82.34** |
| AA(%) | 61.86 | 57.52 | 62.70 | 57.04 | **82.87** | 55.70 | 58.67 | 63.69 | 71.10 |
| Kappa(%) | 70.34 | 70.44 | 74.90 | 63.90 | 70.96 | 67.07 | 67.99 | 73.99 | **80.30** |
| Class1 | 99.15 | 95.03 | **99.92** | 99.52 | 96.10 | 99.81 | 90.36 | 99.60 | 95.44 |
| Class2 | 83.06 | 64.37 | 79.37 | **85.52** | 67.31 | 58.61 | 70.78 | 82.64 | 81.91 |
| Class3 | 74.75 | 68.45 | 72.60 | 68.63 | 70.64 | 63.68 | 70.76 | 75.20 | **80.53** |
| Class4 | 95.26 | 83.36 | 95.77 | 87.44 | 86.54 | 89.23 | 73.63 | **96.12** | 86.60 |
| Class5 | 25.00 | 10.35 | 15.36 | 17.42 | **97.95** | 13.14 | 19.57 | 27.90 | 13.22 |
| Class6 | 82.56 | 90.90 | 90.78 | 89.92 | 70.11 | 87.02 | 73.14 | 92.25 | **96.05** |
| Class7 | 49.53 | 36.12 | 55.29 | 28.11 | **65.67** | 43.53 | 38.05 | 41.79 | 38.83 |
| Class8 | 98.37 | 94.68 | 98.96 | 98.51 | 91.14 | 95.56 | 95.54 | 99.36 | **99.68** |
| Class9 | 59.24 | 67.83 | 60.44 | 47.54 | **93.31** | 46.60 | 60.09 | 63.83 | 73.09 |
| Class10 | 34.14 | 34.83 | 17.76 | 26.58 | **96.22** | 20.03 | 46.41 | 45.87 | 95.36 |
| Class11 | 26.18 | 51.92 | 46.38 | 18.68 | **66.15** | 29.16 | 30.90 | 28.55 | 63.53 |
| Class12 | 4.81 | 7.03 | 8.11 | 7.39 | **93.77** | 4.78 | 30.10 | 7.47 | 11.91 |
| Class13 | 83.79 | 29.94 | 73.56 | 52.53 | **100.00** | 17.15 | 43.93 | 61.95 | 83.17 |
| Class14 | 15.33 | 8.09 | 14.28 | 22.26 | **95.77** | 27.36 | 6.14 | 23.28 | 9.88 |
| Class15 | 71.61 | 63.07 | 76.86 | 68.66 | **80.41** | 68.33 | 76.31 | 70.26 | 63.85 |
| Class16 | 86.90 | 91.04 | 88.67 | 81.77 | 48.65 | 83.18 | 89.21 | 87.21 | **96.62** |
| Class17 | 24.38 | 40.41 | 35.24 | 26.31 | 73.49 | 55.66 | 44.92 | 44.40 | **91.56** |
| Class18 | 99.45 | 98.00 | 99.15 | **99.84** | 98.42 | 99.69 | 96.27 | 98.75 | 98.52 |

Table 14: Detailed classification results on the QY dataset. Best and second-best results are highlighted in bold and underline, respectively.

| Metric | A2S2K | CSCA | DBCT | HiT | MambaLG | SPRLT-Net | SQSFormer | SMFAT | Ours |
|---|---|---|---|---|---|---|---|---|---|
| OA(%) | 73.68 | 56.67 | 69.03 | 70.07 | 74.00 | 68.04 | 68.95 | 73.45 | **75.72** |
| AA(%) | 70.56 | 54.53 | 70.36 | 66.47 | **74.77** | 69.58 | 67.23 | 72.99 | 66.78 |
| Kappa(%) | 66.03 | 44.16 | 60.49 | 61.52 | 66.61 | 59.50 | 60.13 | 66.08 | **68.98** |
| Class1 | 89.53 | 61.19 | 88.48 | 81.49 | **92.94** | 87.53 | 87.23 | 88.29 | 83.24 |
| Class2 | 61.71 | 48.05 | 46.13 | 67.72 | 72.71 | 53.48 | 54.82 | 78.25 | **82.01** |
| Class3 | 10.99 | 8.07 | 10.40 | 6.57 | **61.72** | 6.33 | 9.45 | 7.95 | 10.46 |
| Class4 | 95.23 | 60.05 | **97.55** | 81.85 | 94.01 | 91.44 | 91.00 | 96.52 | 48.17 |
| Class5 | 92.30 | 85.72 | 92.30 | 81.57 | 67.18 | 93.67 | **94.74** | 92.80 | 87.58 |
| Class6 | 73.59 | 64.09 | 87.28 | 79.60 | 60.04 | 85.02 | 66.15 | 74.15 | **89.24** |