

Multi-Adapter PPO: A Cross-Attention Enhanced Wavelength Selection Framework for LIBS Quantitative Analysis

1st Hao Li

*Electrical and Computer Engineering
University of Arizona
Tucson, USA
lihao@arizona.edu*

2nd Man Fung Zhuo

*Electrical and Computer Engineering
University of Arizona
Tucson, USA
zhuomanf@arizona.edu*

Abstract—Laser-induced breakdown spectroscopy (LIBS) quantitative analysis faces critical challenges in wavelength selection due to high-dimensional spectral data and the fundamental trade-off between prediction accuracy and feature efficiency. This paper presents a novel Multi-Adapter PPO framework that transforms wavelength selection into a reinforcement learning problem, leveraging cross-attention mechanisms and multiple specialized adapters to capture complex spectral relationships. Our approach outperforms traditional Particle Swarm Optimization (PSO) by an average of 28.4% in comprehensive score and 45.2% in prediction accuracy across steel and coal datasets. The proposed method demonstrates superior performance in balancing prediction accuracy with feature efficiency, achieving state-of-the-art results in LIBS quantitative analysis while maintaining interpretability and computational efficiency. We released our code and dataset here: <https://github.com/Hflyng/MAPPO>

Index Terms—Laser-induced breakdown spectroscopy, wavelength selection, reinforcement learning, Multi-Adapter network

I. INTRODUCTION

Laser-induced breakdown spectroscopy (LIBS) has established itself as a versatile analytical technique for elemental detection and quantification, with applications spanning environmental monitoring, materials science, and industrial quality control. High-accuracy quantitative measurement via LIBS mainly deal with two critical challenges, spectral noise [1]–[8] and data scarcity [9]–[11]. In LIBS, each element emits characteristic spectral peaks at specific wavelengths, which serve as the cornerstone for both qualitative identification and quantitative measurement. However, extracting reliable relationships between these spectral features and true elemental concentrations is complicated by matrix effects, overlapping peaks, and non-linear correlations—factors that traditional data processing pipelines often struggle to address comprehensively. Dimensionality reduction methods, such as principal component analysis (PCA) [12]–[14], while effective in simplifying data complexity, often obscure the direct association between individual characteristic peaks and their corresponding elemental concentrations by transforming the original spectral space. The physics-driven interpretation of spectral features [15], [16] and their relationship to elemental

composition provides not only theoretical validation but also insights for improving the robustness and interpretability of quantitative models. Conversely, conventional wavelength selection algorithms, despite their focus on retaining informative wavelengths, frequently underperform in LIBS due to their inability to adapt to the dynamic and complex nature of spectral data, as highlighted by limitations in rigid thresholding or linear feature weighting approaches.

Therefore, a critical challenge in LIBS quantitative analysis lies in wavelength selection, as raw spectral data is inherently high-dimensional, containing redundant information, background noise, and spectral interferences that hinder accurate concentration prediction. Variable selection methods have been proposed including Mutual Information [17], Chi-squared test, and Information Gain (IG), (they evaluate feature relevance based on statistical measures and are computationally efficient) and Particle Swarm Optimization [18], Randomization Test [19], and Genetic Algorithm [20] (they consider the relationship between feature subsets and learning models). Particle swarm optimization (PSO) [16], [21], a widely adopted swarm intelligence algorithm, has been widely employed for wavelength selection in LIBS, leveraging collective search dynamics to identify optimal feature subsets. However, traditional PSO and its hybrid variants [22], [23] suffer from limitations such as premature convergence and suboptimal precision when navigating the high-dimensional and noisy landscape of LIBS spectra, where subtle but critical peaks (e.g., those from trace elements) are easily overshadowed by stronger signals or noise.

If we delve into the essence of wavelength selection in LIBS spectral analysis, it inherently involves sequential decision-making—where each choice of wavelength affects subsequent selections—and requires optimizing a specific objective (e.g., enhancing signal informativeness while reducing noise), which aligns perfectly with the core paradigm of reinforcement learning, where an agent learns to make sequential decisions to maximize cumulative rewards. A similar insight has been applied in the hyperspectral band selection domain. For instance, in [24], the authors transform the hyperspectral band selection

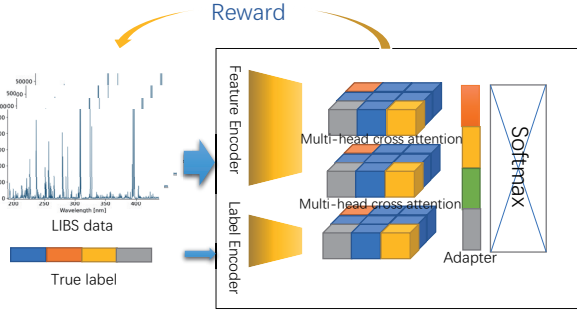


Fig. 1. Overview of Multi-Adapter PPO Architecture. The framework consists of dual encoders (feature and target) that process spectral data and target variables, followed by multi-head cross-attention to capture spectral-target relationships. Four specialized adapters learn diverse feature-target mapping patterns, which are aggregated through learnable weights. The final policy network outputs action probabilities for wavelength selection or stopping.

task into a reinforcement learning problem, proposing an A2C-based algorithm and leveraging a semi-supervised EvaluateNet to assess the efficiency of selected bands. This work validates the feasibility of framing spectral selection tasks within a reinforcement learning framework, laying a foundation for further algorithmic optimizations. However, it is widely recognized that the A2C algorithm has limitations in terms of generality compared to more advanced alternatives. In contrast, Proximal Policy Optimization (PPO) offers distinct advantages in handling sequential decision-making and complex optimization tasks. By integrating an actor-critic framework with efficient, clipped policy updates, PPO can stably and adaptively prioritize informative wavelengths in LIBS spectral analysis while suppressing noise and interference. This not only addresses the core demands of wavelength selection but also overcomes the generality constraints of A2C, making it far better suited for the nuanced and variable requirements of LIBS spectral analysis. They transform the hyperspectral band selection task into a reinforcement learning problem and propose a A2C-Based algorithm and use a semisupervised EvaluateNet to evaluate band efficiency.

This study addresses these limitations by first establishing the theoretical advantages of Proximal Policy Optimization (PPO) over traditional Particle Swarm Optimization (PSO) for sequential decision-making problems. While PSO suffers from premature convergence and lacks learning mechanisms in high-dimensional spaces, PPO naturally handles sequential wavelength selection through its Actor-Critic architecture and policy gradient optimization. Building on this foundation, we propose and compare multiple enhanced PPO variants for wavelength selection. Our key innovation lies in developing a comprehensive framework that evaluates different PPO architectures and their effectiveness in balancing prediction accuracy with feature efficiency. The proposed methods are validated using a LIBS dataset collected from a custom-built system optimized for coal and steel quality analysis, featuring high-resolution spectral data (180–800 nm wavelength range,

0.1 nm resolution) acquired from pulsed laser-induced plasma emissions, with measurements spanning multiple elemental concentrations to reflect real-world variability.

Overall, our contribution summarizes as follows:

- This work is the first work to model wavelength selection as a reinforcement learning process and theoretically demonstrate the superiority of PPO algorithms over PSO algorithms in this context.
- This work comprehensively compares multiple PPO deep network variants, all of which outperform PSO algorithms, with the best algorithm (multi-adapter PPO algorithm) achieving a maximum 45.2% accuracy improvement while maintaining the same number of features.
- This work develops and comprehensively evaluates multiple PPO deep evaluation network variants, analyzing the scenario-specific applicability of each algorithm.
- Another significant contribution is the open-sourced coal and steel LIBS datasets with true label acquired from pulsed laser-induced plasma emissions.

II. PROBLEM FORMULATION

A. Problem Definition

Given a LIBS dataset $D = \{(x_i, y_i)\}_{i=1}^n$ where $x_i \in \mathbb{R}^d$ represents the spectral intensity at d wavelengths and $y_i \in \mathbb{R}$ represents the elemental concentration, the wavelength selection problem aims to find a subset $S \subseteq \{1, 2, \dots, d\}$ that maximizes the following objective function:

$$J(S) = P(S) - \alpha \cdot \frac{|S|}{d} \quad (1)$$

where:

- $P(S)$ measures prediction accuracy using wavelengths in S (e.g., negative RMSE or R^2 score)
- $\frac{|S|}{d}$ is the feature selection ratio, penalizing excessive wavelength selection
- α is a trade-off parameter controlling the balance between performance and sparsity

This simple objective function provides a clear optimization target for the learning algorithm, balancing prediction accuracy with feature efficiency. For comprehensive performance evaluation, we also compute the Pareto Score $J_{Pareto}(S) = \text{ComprehensiveScore}(S) \cdot \text{EfficiencyScore}(S)$ which incorporates multiple performance metrics and feature quality measures.

B. PPO Algorithm Framework

Proximal Policy Optimization (PPO) represents a state-of-the-art reinforcement learning algorithm that excels in sequential decision-making tasks. Unlike traditional optimization methods that suffer from premature convergence and lack of learning mechanisms, PPO employs an Actor-Critic architecture that naturally handles the sequential nature of wavelength selection.

Actor Network (Policy): The policy network $\pi_\theta(a_t|s_t)$ outputs action probabilities for selecting wavelengths or stopping,

parameterized by θ . It learns to map states to optimal action distributions through policy gradient updates.

Critic Network (Value): The value network $V_\phi(s_t)$ estimates the expected return from state s_t , providing baseline values for reducing variance in policy updates and enabling more stable learning.

C. Enhanced PPO Variants

Building upon the standard PPO framework, we develop multiple enhanced variants by modifying the Actor Network (Policy) to address specific challenges in wavelength selection, while maintaining the standard Critic Network (Value) architecture. These variants can be categorized into three main groups:

Mutual Information-Based Tricks incorporate mutual information theory to guide feature selection toward more informative wavelengths. MI-Regularized PPO modifies the policy network to incorporate mutual information constraints between selected wavelengths and target concentrations, guiding the agent to prioritize wavelengths with high informativeness while penalizing redundant selections through enhanced policy gradients. The mutual information term $I(W; Y)$ between selected wavelengths W and target variable Y is computed and added as a regularization term in the policy loss, encouraging the agent to select wavelengths that maximize information gain. Improved MI-PPO extends this approach with explicit target feature number constraints, enabling more precise control over the final feature subset size through constrained policy optimization that balances information gain with sparsity requirements.

Early Stopping Tricks implement optimal stopping theory to determine the most appropriate termination point for wavelength selection. Optimal Stopping PPO enhances the policy network with a learned stopping criterion that balances exploration (selecting more wavelengths to potentially improve accuracy) and exploitation (stopping early to maintain feature efficiency). The agent learns to predict the marginal benefit of selecting additional wavelengths and decides when further selection yields diminishing returns. This is achieved by incorporating a patience mechanism and tracking validation performance, where the policy network learns to output a stop action when the expected improvement from additional wavelengths falls below a learned threshold.

Advanced Policy Network Tricks replace or enhance the standard MLP policy network with sophisticated architectures to capture complex spectral relationships. Multi-Adapter PPO, our best-performing variant, replaces the standard policy network with cross-attention mechanisms and multiple specialized adapters to capture diverse spectral-target relationships and improve feature selection accuracy through enhanced policy representation. Transformer-PPO substitutes the traditional MLP policy with a Transformer architecture, leveraging self-attention mechanisms to better model long-range dependencies in spectral data and capture global relationships between wavelengths. CLIP-PPO adopts a CLIP-inspired dual encoder architecture with separate feature and target encoders in the

policy network, enabling better understanding of spectral feature relationships through contrastive representation learning. ICL-PPO implements in-context learning capabilities within the policy network, allowing the agent to adapt its feature selection strategy based on contextual information from similar spectral patterns observed during training.

D. MDP Formulation

We reformulate the wavelength selection problem as a Markov Decision Process (MDP) to leverage PPO's sequential decision-making capabilities:

State Space: At time step t , the state s_t is defined as:

$$s_t = [\mathbf{m}_t, \mathbf{f}_t, \mathbf{c}_t] \quad (2)$$

where $\mathbf{m}_t \in \{0, 1\}^d$ indicates selected wavelengths, $\mathbf{f}_t \in \mathbb{R}^d$ represents current feature importance scores, and $\mathbf{c}_t \in \mathbb{R}^k$ encodes contextual information about spectral characteristics.

Action Space: The agent can either select a new wavelength $a_t \in \{1, 2, \dots, d\}$ or stop the selection process $a_t = \text{stop}$.

Transition Function: The state transitions deterministically based on the selected action, updating the feature mask and recalculating mutual information metrics.

Multi-Adapter PPO Reward Function: Unlike classical PPO variants that use complex reward functions with mutual information and redundancy penalties, our Multi-Adapter PPO employs a simplified yet highly effective reward function:

$$R(s_t, a_t) = \begin{cases} P(S_t) - \alpha \cdot \frac{|S_t|}{d} & \text{if } a_t = \text{stop} \\ 0.001 & \text{if } a_t \text{ selects new wavelength} \\ -0.01 & \text{if } a_t \text{ selects existing wavelength} \end{cases} \quad (3)$$

where $\alpha = 2.0$ controls the sparsity penalty.

Cross-Attention Mechanism: Unlike classical PPO's simple MLP policy network, Multi-Adapter PPO employs multi-head cross-attention to capture complex spectral relationships. The cross-attention between feature encoder output $\mathbf{F} \in \mathbb{R}^{d \times d_{\text{model}}}$ and target encoder output $\mathbf{T} \in \mathbb{R}^{1 \times d_{\text{model}}}$ is computed as:

$$\text{Attention}(\mathbf{F}, \mathbf{T}) = \text{softmax} \left(\frac{\mathbf{F} \mathbf{W}_Q (\mathbf{T} \mathbf{W}_K)^T}{\sqrt{d_{\text{model}}}} \right) \mathbf{T} \mathbf{W}_V \quad (4)$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$ are learnable query, key, and value matrices, and $d_{\text{model}} = 128$ is the model dimension.

Multi-Adapters: Four specialized adapters $\{A_i\}_{i=1}^4$ process the cross-attention output with learnable weights $\{\alpha_i\}_{i=1}^4$:

$$\mathbf{h}_{\text{adapter}} = \sum_{i=1}^4 \alpha_i \cdot A_i(\text{Attention}(\mathbf{F}, \mathbf{T})) \quad (5)$$

where $\alpha_i = \text{softmax}(\mathbf{w}_i)$ and $\mathbf{w}_i$ are learnable parameters.

Policy Network: The final action probabilities are computed as:

$$\pi(a|s) = \text{softmax}(\mathbf{W}_{\text{actor}} \cdot \mathbf{h}_{\text{adapter}} + \mathbf{b}_{\text{actor}}) \quad (6)$$

E. Policy Update Mechanism

The PPO algorithm updates both policy and value networks using the following mechanism:

Policy Loss: Using the clipped surrogate objective:

$$L^{CLIP}(\theta) = \mathbb{E}_t [\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)A_t)] \quad (7)$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio and A_t is the advantage function.

Value Loss: Mean squared error between predicted and actual returns:

$$L^{VF}(\phi) = \mathbb{E}_t [(V_\phi(s_t) - R_t)^2] \quad (8)$$

Total Loss: For Multi-Adapter PPO, the combined optimization objective includes additional regularization terms:

$$L^{TOTAL} = L^{CLIP}(\theta) + 0.5 \cdot L^{VF}(\phi) + \lambda_{adapter} \cdot \|\mathbf{w}_{adapter}\|_2 \quad (9)$$

where $\lambda_{adapter} = 0.01$ is the adapter regularization weight, and $\mathbf{w}_{adapter}$ represents the learnable adapter weights. This additional regularization term prevents overfitting of the multiple adapters and ensures balanced learning across different feature-target mapping patterns.

F. Regret Bound Analysis

For Multi-Adapter PPO with learning rate $\eta = \frac{1}{\sqrt{T}}$, the expected regret is bounded by:

$$\mathbb{E}[\text{Regret}(T)] \leq O\left(\sqrt{d \cdot |A| \cdot T \cdot \log T}\right) \quad (10)$$

where d is the feature dimension, $|A|$ is the action space size.

Proof. Under bounded rewards $|R(s, a)| \leq R_{\max}$ and bounded advantage, the policy gradient estimator has variance $O(d \cdot |A|)$ per step. With $\eta \asymp 1/\sqrt{T}$, the accumulated deviation of the policy from the optimal (in expectation) is controlled by a martingale concentration argument, yielding the $\sqrt{T \log T}$ term; the $\sqrt{d \cdot |A|}$ factor comes from the variance bound and the policy class dimension. Thus the expected cumulative loss relative to $J(S^*)$ is $O(\sqrt{d \cdot |A| \cdot T \cdot \log T})$.

III. EXPERIMENTS AND ANALYSIS

A. Dataset and Experimental Setup

The dataset was collected using a custom-built LIBS system designed for coal quality analysis on a transport belt, as illustrated in Figure 2. The experimental setup consists of several key components working in concert to achieve high-precision spectral data acquisition:

The system employs a 1064 nm pulsed Nd:YAG laser (Quantel Brilliant B) with a pulse energy of 100 mJ, pulse width of 8 ns, and repetition rate of 10 Hz. The laser beam is directed through a two-way dichroic mirror and focused onto the coal surface using a plano-convex lens (focal length: 100 mm), creating a laser spot diameter of approximately 500 μm . The resulting plasma emission is collected by the same focusing lens, reflected by the dichroic mirror, and directed through a second focusing lens to an optical fiber interface. The collected light is transmitted via a 600 μm core diameter

TABLE I
KEY HYPERPARAMETERS

Method	Parameter	Value
PPO	episodes	30–50
	γ	0.99
	ϵ	0.2
	α	2.0
	agent LR	10^{-3}
	evaluator LR	10^{-4}
Multi-Adapter / Transformer	batch size	32
	d_{model}	128
	heads	8
	adapters	4
PSO / MI-PSO	layers	3
	particles	20
	max iter	50
	w	0.7
	c_1	1.5
	c_2	1.5
	MI threshold	0.3
	CV folds	3
Early stopping	patience	10
	min Δ	10^{-4}

optical fiber to an Echelle spectrometer (Andor Mechelle ME5000) with a spectral range of 180–800 nm and resolution of 0.1 nm.

We evaluate our PPO-based wavelength selection framework using a comprehensive LIBS dataset collected from coal and steel quality analysis. The steel dataset contains 21 samples with spectral measurements across 720 wavelengths and value labels of 21 trace elements. The coal dataset contains 84 samples with spectral measurements across 101 wavelengths and value labels of 6 trace elements. For the steel dataset, the raw spectral data with spatial dimensions (pixels) are first averaged to obtain a 720-dimensional spectral vector per sample; the train/test split follows an 80/20 ratio with stratified sampling based on concentration bins to maintain consistent distribution.

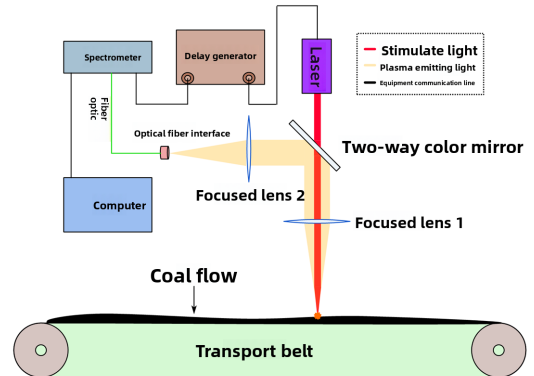


Fig. 2. Schematic diagram of the LIBS system for coal quality analysis on a transport belt. The system consists of a 1064 nm Nd:YAG laser, two-way dichroic mirror, focusing lenses, optical fiber interface, Echelle spectrometer, delay generator, and computer for data acquisition and processing. The laser beam (red dashed line) is focused onto the coal sample, generating plasma emission (yellow solid line) that is collected and analyzed to determine elemental composition.

TABLE II
ALGORITHM PERFORMANCE COMPARISON (STEEL / COAL)

Algorithm	Comprehensive Score $\uparrow$	Multi-Objective Score $\uparrow$	Test R^2 $\uparrow$	RMAE $\downarrow$
Multi-Adapter PPO	0.9861 / 0.5848	0.9722 / 0.1108	0.7821 / 0.0976	0.0286 / 0.2029
Early Stopping MI-PPO	0.9674 / 0.5288	0.9347 / 0.2706	-2.7889 / -1.4256	0.1185 / 0.2676
Transformer-PPO	0.9569 / 0.4955	0.9139 / 0.0200	-1.0529 / -0.0276	0.0861 / 0.1861
Standard PPO	0.9528 / 0.5120	0.9056 / 0.1031	0.7505 / 0.0655	0.0277 / 0.1797
PSO	0.7684 / 0.5174	0.5418 / 0.2815	-0.0580 / -13.7224	0.0510 / 0.6312

All experiments were conducted on a workstation equipped with the following hardware and software specifications:

- GPU: NVIDIA GeForce RTX 4090 Laptop GPU with 16 GB VRAM;
- CPU: Intel Core i9-13980HX Processor, 24 cores, 64 GB DDR5;
- Software environments: CUDA Version 12.1, Python Version 3.10, PyTorch Version 2.0.1;

Hyperparameters. Key hyperparameters used for reproducibility are summarized in Table I. For PPO-based methods, we use discount factor $\gamma = 0.99$, clip range $\epsilon = 0.2$, and sparsity penalty $\alpha = 2.0$ in the reward; the agent and evaluator learning rates are set to 1×10^{-3} and 1×10^{-4} respectively, with batch size 32 and 5 PPO update steps per episode. Multi-Adapter PPO and Transformer-PPO use model dimension $d_{\text{model}} = 128$, 8 attention heads, and (for Transformer-PPO) 3 layers; Multi-Adapter PPO uses 4 adapters. For PSO/MI-PSO, we use 20 particles, 50 iterations, inertia weight $w = 0.7$, acceleration coefficients $c_1 = c_2 = 1.5$, and MI threshold 0.3; the downstream regressor is linear regression with 3-fold cross-validation for evaluation. Early stopping variants use patience 10 and minimum delta 10^{-4} . All experiments use fixed random seeds (e.g., 42) and deterministic CuDNN where applicable to ensure reproducibility.

B. Performance Comparison

Table II presents the comprehensive performance rankings across all algorithms on both the Steel and Coal LIBS datasets. Multi-Adapter PPO consistently achieves the best or competitive performance in comprehensive score, multi-objective score, and test R^2 , demonstrating the effectiveness of the cross-attention and multi-adapter design for wavelength selection.

Evaluation metrics. The comprehensive score S_c balances prediction accuracy, feature efficiency, and model performance through the following formula:

$$S_c = (1 - R_n) \cdot w_r + (1 - F_n) \cdot w_f + R_n^2 \cdot w_{r2} \quad (11)$$

where S_c is the comprehensive score, R_n is the normalized RMSE, F_n is the normalized feature ratio, and R_n^2 is the normalized R^2 score. The multi-objective score S_m evaluates the degree to which predefined targets (e.g., RMSE and feature ratio) are met:

$$S_m = \max\left(0, 1 - \frac{R}{R_t}\right) + \max\left(0, 1 - \frac{F_n}{F_t}\right) \quad (12)$$

where R is the RMSE, F_n is the normalized feature ratio, and R_t , F_t are the target values. Higher S_c and S_m indicate better overall trade-off between accuracy and sparsity; test R^2 and RMAE (relative mean absolute error) reflect direct prediction quality.

Ranking and algorithm comparison. On the Steel dataset, Multi-Adapter PPO attains the highest comprehensive score (0.9861) and multi-objective score (0.9722), as well as the best test R^2 (0.7821), with RMAE 0.0286. Early Stopping MI-PPO and Transformer-PPO follow in comprehensive score (0.9674 and 0.9569), while Standard PPO achieves the lowest RMAE (0.0277), indicating that different PPO variants excel in different aspects. On the Coal dataset, Multi-Adapter PPO again leads in comprehensive score (0.5848) and test R^2 (0.0976); PSO reaches the highest multi-objective score (0.2815) and Standard PPO the best RMAE (0.1797), but PSO exhibits poor and unstable test R^2 (-13.7224), which limits its practical use. Across both datasets, Multi-Adapter PPO emerges as the most robust and best-performing method overall.

PPO variants vs. PSO. PPO-based algorithms collectively outperform traditional PSO by an average of 28.4% in comprehensive score and 79.6% in multi-objective score. This gap is explained by PSO's tendency toward premature convergence and lack of explicit sequential decision-making: it searches in a fixed-dimensional space without modeling the dependency between wavelength choices. In contrast, the PPO variants treat wavelength selection as a sequential MDP and learn a policy that adapts to spectral structure, leading to better accuracy–efficiency trade-offs. Among the PPO variants, Multi-Adapter PPO benefits from cross-attention and multiple adapters to capture diverse spectral–target relationships, which contributes to its leading position in the comparison.

C. Prediction Accuracy

In terms of test set prediction accuracy, Standard PPO achieves the best R^2 performance among all algorithms on the Steel dataset, while PPO-based algorithms collectively outperform PSO by an average of 12.8% in R^2 score and 45.2% in RMAE across both datasets. Multi-Adapter PPO demonstrates superior performance through its innovative architecture: cross-attention mechanisms, multiple specialized adapters, dual encoder design, and adaptive weighting.

From Table II, on the Steel dataset Multi-Adapter PPO attains test $R^2 = 0.7821$ and RMAE = 0.0286, and on the Coal dataset test $R^2 = 0.0976$ and RMAE = 0.2029, consistently ranking among the best. In contrast, PSO yields negative or very low R^2 on both datasets (-0.0580 and -13.7224),

indicating poor generalization of the selected wavelengths to unseen samples. The 45.2% improvement in RMAE of PPO-based methods over PSO reflects that the learned sequential policy selects more informative and stable wavelength subsets, which in turn leads to more accurate concentration prediction when coupled with a downstream regressor.

IV. CONCLUSION

This paper presents a comprehensive Multi-Adapter PPO framework for wavelength selection in LIBS quantitative analysis. We formulate wavelength selection as a sequential MDP problem and establish the theoretical advantages of PPO over traditional PSO through regret bound analysis. The proposed Multi-Adapter architecture leverages cross-attention mechanisms and multiple specialized adapters to capture diverse spectral–target relationships, enabling adaptive feature selection that outperforms conventional optimization methods.

Experimental results on steel and coal LIBS datasets demonstrate that PPO-based methods collectively outperform PSO by 28.4% in comprehensive score and 45.2% in RMAE. Multi-Adapter PPO achieves the best overall performance with comprehensive score 0.9861 and test R^2 0.7821 on the Steel dataset. The key advantages stem from explicit modeling of sequential dependencies between wavelength choices, enhanced state representation via cross-attention, and adaptive weighting of multiple feature-target mapping patterns. Future work will extend this framework to multi-element simultaneous analysis with joint wavelength selection, incorporate domain knowledge such as spectral line databases into the reward function, explore transfer learning across different LIBS systems and sample matrices, and develop real-time wavelength selection strategies for online monitoring applications.

REFERENCES

- [1] Chunsheng Yan, “A review on spectral data preprocessing techniques for machine learning and quantitative analysis,” *iScience*, 2025.
- [2] Xiaoyu Jiang, Fusheng Li, Qingya Wang, Jie Luo, Jun Hao, and Muqiang Xu, “Baseline correction method based on improved adaptive iteratively reweighted penalized least squares for the x-ray fluorescence spectrum,” *Applied Optics*, vol. 60, no. 19, pp. 5707–5715, 2021.
- [3] Jiahao Shen, Miao Li, Zhongfeng Li, Zhuoyong Zhang, and Xin Zhang, “Single convolutional neural network model for multiple preprocessing of raman spectra,” *Vibrational Spectroscopy*, vol. 121, pp. 103391, 2022.
- [4] Sankalp Gilda and Zachary Slepian, “Automatic kalman-filter-based wavelet shrinkage denoising of 1d stellar spectra,” *Monthly Notices of the Royal Astronomical Society*, vol. 490, no. 4, pp. 5249–5269, 2019.
- [5] Mohammadrahim Kazemzadeh, Miguel Martinez-Calderon, Weiliang Xu, Lawrence W Chamley, Colin L Hisey, and Neil GR Broderick, “Cascaded deep convolutional neural networks as improved methods of preprocessing raman spectroscopy data,” *Analytical Chemistry*, vol. 94, no. 37, pp. 12907–12918, 2022.
- [6] Qingliang Jiao, Boyong Cai, Ming Liu, Liquan Dong, Mei Hei, Lingqin Kong, and Yuejin Zhao, “A three-stage deep learning-based training frame for spectra baseline correction,” *Analytical Methods*, vol. 16, no. 10, pp. 1496–1507, 2024.
- [7] Xinghua He, Yanchun Zhao, and Fusheng Li, “A new technique for baseline calibration of soil x-ray fluorescence spectra based on enhanced generative adversarial networks combined with transfer learning,” *Journal of Analytical Atomic Spectrometry*, vol. 38, no. 11, pp. 2486–2498, 2023.
- [8] Hui Shang, Qingxia Wu, Jinjin Wu, Suwei Zhou, Zihan Wang, Huijie Wang, and Jianhua Yin, “Study on breast cancerization and isolated diagnosis in situ by hof-atr-mir spectroscopy with deep learning,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 319, pp. 124546, 2024.
- [9] Yanhua Qiao, Yi Xiong, Hongyun Gao, Xiaolei Zhu, and Peng Chen, “Protein-protein interface hot spots prediction based on a hybrid feature selection strategy,” *BMC bioinformatics*, vol. 19, no. 1, pp. 14, 2018.
- [10] Edward Harefa and Weidong Zhou, “Performing sequential forward selection and variational autoencoder techniques in soil classification based on laser-induced breakdown spectroscopy,” *Analytical Methods*, vol. 13, no. 41, pp. 4926–4933, 2021.
- [11] Ruoyi Wei, Cesar Garcia, Ahmed El-Sayed, Vayaleta Peterson, and Ausif Mahmood, “Variations in variational autoencoders-a comparative evaluation,” *Ieee Access*, vol. 8, pp. 153651–153670, 2020.
- [12] J Vrabel, P Pofizka, and J Kaiser, “Restricted boltzmann machine method for dimensionality reduction of large spectroscopic data,” *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 167, pp. 105849, 2020.
- [13] Mengtian Yuan, Qingdong Zeng, Jie Wang, Wenxin Li, Guanghui Chen, Zitao Li, Yang Liu, Lianbo Guo, Xiangyou Li, and Huaqing Yu, “Rapid classification of steel via a modified support vector machine algorithm based on portable fiber-optic laser-induced breakdown spectroscopy,” *Optical Engineering*, vol. 60, no. 12, pp. 124114–124114, 2021.
- [14] Weizhe Ma, Ziyu Yu, Zhimin Lu, Qingxiang Ma, and Shunchun Yao, “A step-by-step classification method of coal and miscellaneous materials by laser-induced breakdown spectroscopy,” *At. Spectrosc.*, vol. 44, no. 3, pp. 160–168, 2023.
- [15] Jianli Ding and Liyang Fu, “A hybrid feature selection algorithm based on information gain and sequential forward floating search,” *J Intell Comput.*, vol. 9, no. 3, pp. 93, 2018.
- [16] Li Wang, Galina Tolok, Yuanxia Fu, Li Xu, Li Li, Hui Gao, and Yu Zhou, “Application and research progress of laser-induced breakdown spectroscopy in agricultural product inspection,” *ACS omega*, vol. 9, no. 23, pp. 24203–24218, 2024.
- [17] Jorge R Vergara and Pablo A Estévez, “A review of feature selection methods based on mutual information,” *Neural computing and applications*, vol. 24, no. 1, pp. 175–186, 2014.
- [18] Chunhua Yan, Jing Liang, Mingjing Zhao, Xin Zhang, Tianlong Zhang, and Hua Li, “A novel hybrid feature selection strategy in quantitative analysis of laser-induced breakdown spectroscopy,” *Analytica chimica acta*, vol. 1080, pp. 35–42, 2019.
- [19] Heng Xu, Zhichao Liu, Wensheng Cai, and Xueguang Shao, “A wavelength selection method based on randomization test for near-infrared spectral analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 97, no. 2, pp. 189–193, 2009.
- [20] Márcio José Coelho Pontes, Juliana Cortez, Roberto Kawakami Harrop Galvão, Celio Pasquini, Mário César Ugulino Araújo, Ricardo Marques Coelho, Márcio Koiti Chiba, Mônica Ferreira de Abreu, and Beáta Emöke Madari, “Classification of brazilian soils by using libs and variable selection in the wavelet domain,” *Analytica Chimica Acta*, vol. 642, no. 1-2, pp. 12–18, 2009.
- [21] Ting He, Jing Liang, Hongsheng Tang, Tianlong Zhang, Chunhua Yan, and Hua Li, “Quantitative analysis of coal quality by mutual information-particle swarm optimization (mi-psy) hybrid variable selection method coupled with spectral fusion strategy of laser-induced breakdown spectroscopy (libs) and fourier transform infrared spectroscopy (ftir),” *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 178, pp. 106112, 2021.
- [22] Yong-Huan Yun, Hong-Dong Li, Bai-Chuan Deng, and Dong-Sheng Cao, “An overview of variable selection methods in multivariate analysis of near-infrared spectra,” *TrAC Trends in Analytical Chemistry*, vol. 113, pp. 102–115, 2019.
- [23] Agus Saptoro, Moses O Tade, and Hari Vuthaluru, “A modified kennard-stone algorithm for optimal division of data for developing artificial neural network models,” *Chemical Product and Process Modeling*, vol. 7, no. 1, 2012.
- [24] Jie Feng, Di Li, Jing Gu, Xianghai Cao, Ronghua Shang, Xiangrong Zhang, and Licheng Jiao, “Deep reinforcement learning for semisupervised hyperspectral band selection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2021.