Journal of Hazardous Materials 505 (2026) 141490

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

A multi-view linearly constrained tabular prior-data fitted network (MLC-TabPFN) model for arsenic quantification in hyperaccumulator using LIBS

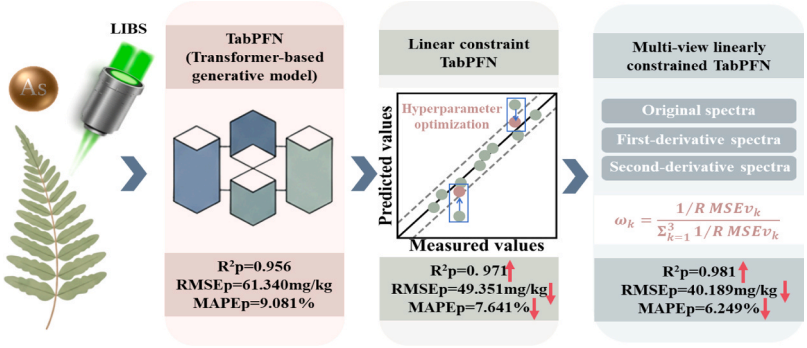
Liuye Cao [a], Jing Huang [b], Lei Feng [a], Dongxing Guan [c], Zhihua Dai [d], Fei Liu [a,e,*]

[a] State Key Laboratory for Vegetation Structure, Function and Construction (VegLab), College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China
[b] College of Agricultural Science and Engineering, Hohai University, No.8 West Focheng Road, Nanjing 211100, China
[c] State Key Laboratory of Soil Pollution Control and Safety, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China
[d] School of Ecology and Environment, Anhui Normal University, Wuhu 241002, China
[e] Zhejiang Key Laboratory of Agricultural Remote Sensing and Information Technology, Zhejiang University, Hangzhou 310058, China

HIGHLIGHTS

- A novel MLC-TabPFN model is proposed for quantitative LIBS analysis.
- The baseline TabPFN model outperforms traditional algorithms like PLSR
- A knowledge-driven linear constraint improves model accuracy and plausibility.
- Multi-view fusion of spectral data significantly boosts predictive performance.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Pteris vittata
Arsenic
Laser-induced breakdown spectroscopy (LIBS)
Tabular prior-data fitted network (TabPFN)
Multi-view

ABSTRACT

Rapid and accurate quantification of arsenic (As) in the hyperaccumulator *Pteris vittata* is vital for evaluating the efficacy of phytoremediation. While Laser-Induced Breakdown Spectroscopy (LIBS) offers a promising rapid analysis solution, its precision is often compromised by complex biological matrix effects and the limitations of conventional chemometric models. To address these challenges, this study proposes a novel quantitative model, the multi-view linearly constrained tabular prior-data fitted network (MLC-TabPFN), which synergistically combines data-driven learning with knowledge-driven principles. Experimental results first established that the baseline TabPFN model surpassed traditional algorithms like PLSR, achieving a higher prediction coefficient of determination (R^2p) of 0.956, compared to 0.933 for PLSR. Building on this strong foundation, two key innovations were introduced. First, a linear constraint (LC) module was integrated, which incorporates the physical prior of a positive correlation between spectral intensity and concentration. This knowledge-driven enhancement improved the model's plausibility and accuracy, significantly reducing the mean absolute percentage error

* Corresponding author at: State Key Laboratory for Vegetation Structure, Function and Construction (VegLab), College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China.
E-mail address: fliu@zju.edu.cn (F. Liu).

https://doi.org/10.1016/j.jhazmat.2026.141490
Received 30 October 2025; Received in revised form 30 December 2025; Accepted 14 February 2026
Available online 16 February 2026

(MAPEp) from 8.537 % to 7.641 % while maintaining a high coefficient of determination ($R^2p = 0.971$) on the optimal second-derivative spectra. Second, a multi-view fusion module was implemented to integrate complementary information from original, first-, and second-derivative spectra. This final step significantly boosted performance, elevating the R^2p to 0.981, while substantially lowering the RMSEp to 40.189 mg/kg and the MAPEp to 6.249 %. This study validates that the proposed framework provides a superior solution for high-precision LIBS analysis in complex biological matrices, demonstrating a powerful synergy between advanced machine learning and physical domain knowledge.

1. Introduction

Arsenic (As) is a highly toxic and carcinogenic heavy metal, and its widespread presence in soil and water represents a major global challenge that requires urgent attention [1]. Long term exposure to arsenic contaminated environments poses a severe threat to both ecosystems and human health, underscoring the critical need for developing efficient remediation technologies [2]. Among various approaches, phytoremediation has gained considerable attention in recent years as a green, cost-effective, and sustainable solution. Within the group of known hyperaccumulators, the *Pteris vittata* is widely recognized as the most promising species for the remediation of arsenic contaminated sites [3]. Research has demonstrated that this plant not only exhibits exceptional tolerance to arsenic but also accumulates substantial amounts of it in its above ground biomass. The arsenic concentration in its fronds can reach levels over one hundred times greater than that in the surrounding soil [4]. This unique capability makes *P. vittata* a highly valuable candidate for ecological restoration projects in arsenic polluted areas. Consequently, the ability to rapidly and accurately monitor arsenic accumulation in *P. vittata* is essential for evaluating the efficacy of remediation efforts, optimizing field management strategies, and understanding the mechanisms of arsenic translocation and transformation within the plant.

For the determination of arsenic content, inductively coupled plasma mass spectrometry (ICP-MS) and atomic absorption spectrometry (AAS) are widely used as traditional techniques. These methods offer high precision and sensitivity, enabling quantitative analysis at trace and even ultra trace levels [5]. For instance, studies have reported that the limit of detection for ICP-MS can be as low as 0.022 mg/L, allowing for the highly sensitive measurement of arsenic [6]. As a result, they are widely considered the gold standard in fields like environmental monitoring and food safety. However, they also present notable limitations, including complex sample preparation that often requires digestion with strong acids, relatively long analysis times, and difficulties in performing rapid, in situ, or on-site analysis [7]. These drawbacks severely hamper their application in phytoremediation for large scale field monitoring and high throughput sample screening. Therefore, developing a technology that enables rapid, in situ, detection of arsenic is critical for advancing the widespread implementation and real time assessment of phytoremediation projects.

Laser-induced breakdown spectroscopy (LIBS) has emerged as a novel spectroscopic technique that requires only a small amount of sample, involves minimal or no pretreatment, enables simultaneous multi-element rapid analysis, and offers the potential for in situ measurements [8]. The technique operates by focusing a high energy laser pulse onto a sample to generate a plasma; the characteristic light emitted as the plasma cools is then collected for qualitative and quantitative elemental analysis [9]. LIBS has been widely applied to heavy metal detection in soils, water, and plant samples. However, despite its potential, the application of LIBS to complex biological matrices remains challenging. Strong matrix effects, significant spectral interferences, and poor signal stability often compromise the analytical precision and robustness of quantitative measurements [10].

To address these challenges, researchers have widely incorporated chemometric methods, including machine learning algorithms such as partial least squares regression (PLSR), random forest (RF), and extreme gradient boosting (XGBoost). For instance, Peng et al. applied a full spectra PLSR model to detect multiple mineral elements in rice fronds under chromium stress. The model achieved Rp (coefficient of determination for the prediction set) values of 0.9637, 0.8712, 0.8933, 0.8969, and 0.9219 for Ca, Fe, K, Mg, and Na, with corresponding RMSEp (Root Mean Squared Error for the prediction set) values of 281.84, 14.78, 1192.47, 436.07, and 293.82 mg/kg, respectively [8]. Similarly, Xie et al. employed transfer learning from rice husks to improve an XGBoost model for cadmium detection, which substantially reduced prediction error; for brown rice, the prediction set R^2 (coefficient of determination) reached 0.9743, with an RMSEp of 0.0039 mg/kg [11]. These findings indicate that machine learning has shown clear advantages in LIBS-based quantitative analysis. However, its practical application still faces two interrelated core challenges. First, on the model side, many complex algorithms exhibit a clear black-box nature, lacking sufficient interpretability. This makes it generate predictions that contradict fundamental spectroscopic principles in some cases [12]. Second, on the data side, the selection and optimization of data preprocessing methods is critical to the reliability of LIBS quantitative analysis. Traditional methods such as standard normal variate (SNV), multiplicative scatter correction (MSC), and first-/second-derivative transformations have played an important role in correcting scattering, baseline shifts, and noise. However, there is no universal gold standard, and their selection often relies on empirical trial and error [13]. More critically, relying on a single preprocessing view fails to exploit the complementary information available across different representations (e.g., original spectra, first-/second-derivative spectra), potentially leading to information loss or noise amplification [14].

To address these challenges, we propose the multi-view linearly constrained tabular prior-data fitted network (MLC-TabPFN). Unlike traditional approaches, this model integrates data-driven learning with knowledge-driven constraints. Specifically, the model uses TabPFN as the foundation and employs a linear constraint module (LC module) to enforce physical priors, ensuring predictions respect the positive correlation between spectral intensity and concentration, thereby resolving the black-box issue. Simultaneously, it adopts a multi-view strategy to fuse original spectra with their first- and second-derivative information, effectively utilizing complementary features. This multi view, physically constrained approach offers a reliable solution for the rapid and accurate determination of As content in *P. vittata*.

2. Materials and methods

2.1. Sample preparation and As measurement

The As hyperaccumulator *P. vittata* was employed in this study due to its remarkable ability to tolerate and sequester As. A total of 30 seedlings of similar size (approximately 10 cm in height) were selected. Before the As treatment, the plants were kept in a greenhouse for two weeks. The conditions were a 14 h light and 10 h dark cycle, day and night temperatures of 26 °C and 20 °C, relative humidity of 60–70 %, and light intensity of about 350 μmol m^{-2} s^{-1} . The plants were grown hydroponically in 1 L containers with 0.2 × Hoagland nutrient solution. After acclimation, fresh solutions were prepared by adding sodium hydrogen arsenate heptahydrate ($Na_2HAsO_4{\cdot}7H_2O$) to the nutrient medium to provide different As concentrations. Sodium nitroprusside (SNP) was

added as a nitrogen source. Each treatment had three replicates. The plants were exposed for 14 days. At harvest, fronds were collected and air dried in a clean room. The dried tissues were ground into fine powder with a laboratory mill.

For As analysis, 0.50 g of powder was digested with nitric acid (HNO_3) and hydrogen peroxide (H_2O_2) using a graphite block digestion system (Environmental Express, Mt. Pleasant, SC, USA). The digests were filtered with a 0.45 μm cellulose nitrate membrane. The clear solutions were analyzed using an inductively coupled plasma mass spectrometer (ICP-MS, NexION 300X, Perkin Elmer, Waltham, MA, USA). Analytical precision was verified against certified plant reference materials, and calibration curves consistently yielded $R^2 > 0.999$.

2.2. LIBS data acquisition

For LIBS, 0.20 g of powder was pressed into pellets with a hydraulic press. The pellets had a diameter of 13 mm and a thickness of about 1.5 mm. Each biological sample was prepared in triplicate. In total, 90 pellets were obtained for LIBS analysis. Pressing ensured that the pellet surfaces were flat and uniform for measurement.

The LIBS spectra of *P. vittata* were obtained using a laboratory-built system. A schematic of the setup is provided in Fig. 1. A Q-switched Nd:YAG laser operating at 532 nm with a pulse duration of 8 ns was used as the excitation source. The laser output was adjusted to 90 mJ per pulse at a repetition rate of 1 Hz. A plano-convex lens with a focal length of 100 mm was applied to focus the laser beam approximately 2 mm beneath the pellet surface, which produced a stable plasma. The emitted plasma light was collected and dispersed with a spectrograph (SR-500i-A-R, Andor, Belfast, UK) covering the spectral range of 220–240 nm, with a resolution of 0.03 nm. The signals were recorded by an intensified charge-coupled device (ICCD, DH334T-18F-03, Andor, Belfast, UK). The ICCD was triggered with a delay of 2 μs and a gate width of 30 μs to reduce background interference. For each pellet, laser ablation was performed at 16 positions arranged in a 4 × 4 grid. Each position was irradiated five times, resulting in a total of 80 spectra per sample. The spectra from all ablation events were then averaged to obtain a representative spectrum for each sample [15]. These 80 spectra were averaged into a single representative spectrum to reduce the effects of potential inhomogeneity across the pellet surface. Despite homogenizing and pressing the *P. vittata* samples, there may still be variations at the micron-scale laser spot, and averaging helps to capture the overall spectral characteristics, providing a more reliable and consistent input for model training.

2.3. Data processing and modeling

In this study, multivariate models and MLC-TabPFN were used for modeling. The dataset was divided into calibration, validation, and prediction sets at a ratio of 3:1:1. Crucially, this splitting was performed at the sample level rather than the spectrum level to ensure that all spectra from a given sample reside exclusively in one subset. This procedure guarantees the independence of the test set and effectively prevents data leakage. Model performance was then evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE) [16]. For clarity, R^2, RMSE, and MAPE were calculated separately for the calibration set (R^2c, RMSEc, MAPEc), validation set (R^2v, RMSEv, MAPEv), and prediction set (R^2p, RMSEp, MAPEp).

2.3.1. Multivariate models

PLSR projects predictors and responses into a latent space that maximizes their covariance, which is suitable for collinear spectral data [8]; in this study, the number of latent components was set to 15 to achieve the lowest RMSEp. RF aggregates multiple decision trees built on bootstrap samples to reduce variance and enhance stability [17]; here, 100 trees were constructed with default splitting rules. XGBoost builds trees sequentially by minimizing a regularized loss function through gradient boosting, allowing the capture of complex nonlinear relations [11]; it was configured with 50 estimators, a learning rate of 0.1, and a maximum depth of 4. Categorical boosting (CatBoost) improves boosting with ordered boosting and symmetric trees to mitigate overfitting and prediction shift [18]; it was run with 100 iterations, depth 6, and a learning rate of 0.03. Extreme learning machine (ELM) assigns random weights and biases to the hidden layer and then computes output weights analytically by least squares, achieving efficient nonlinear modeling [19]; it was implemented as a single hidden layer neural network with 10 hidden neurons.

Tabular prior-data fitted network (TabPFN) is a transformer-based model tailored for tabular data prediction [20]. It stands out by utilizing a large set of synthetic datasets generated via Structural Causal Models (SCMs) for pre-training, which allows the model to learn how to predict target values based on input features without relying on extensive feature engineering or manual data preprocessing. TabPFN employs a dual-path attention mechanism, where each cell first attends to features within its row (intra-sample attention) and then to the same feature across all samples (intra-feature attention), capturing complex dependencies in tabular data. The model uses in-context learning (ICL) to make predictions in a single forward pass, adapting quickly to new datasets without needing retraining. This makes it highly efficient, particularly in scenarios with smaller datasets, where retraining would typically be time-consuming. Its ability to handle missing values, noise, and other data challenges makes it a powerful tool for real-world tabular data tasks. In this study, the original LIBS spectra were transformed into matrices from the original spectra, the first-order derivative spectra, and

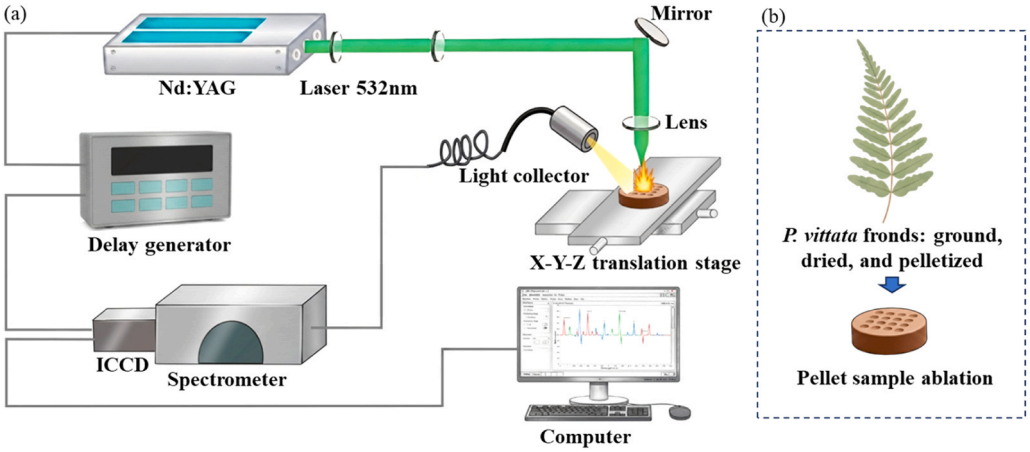


Fig. 1. (a) Schematic diagrams of the LIBS experimental setup. (b) Schematic diagrams of sample pellet and ablation spot arrangement.

the second-order derivative spectra. The TabPFN model was applied under each spectral view, and its performance was evaluated. For clarity, TabPFN-0, TabPFN-1, and TabPFN-2 denote the TabPFN models constructed under the original spectra view, first-derivative view, and second-derivative view, respectively. Traditional multivariate models, including TabPFN, served as baseline references for comparison with the proposed MLC-TabPFN framework.

2.3.2. MLC-TabPFN

To enhance the performance of the TabPFN model, we incorporate spectral physical prior knowledge into the model, termed the linearly constrained TabPFN (LC-TabPFN). Under the local thermodynamic equilibrium (LTE) condition, the emission intensity of element specific spectral lines in LIBS plasma theoretically exhibits a positive correlation with the particle number density of the element in the plasma (and consequently with the atomic concentration in the sample) [21]. The core mechanism of the LC module is based on the assumption that, although the relationship between LIBS spectral intensity and elemental concentration is highly nonlinear due to factors such as matrix effects and self-absorption, a potential linear association can still be preserved within characteristic spectral regions. The detailed procedure of the LC module is summarized as follows.

Select the wavelength features that show a strong correlation with As concentration. Calculate the Pearson correlation coefficient between each wavelength and the measured concentration, then pick the top n wavelengths with the highest correlations as key features. The multiple linear regression model was constructed using the selected key wavelengths, as defined in Eq. (1).

$$\widehat{Y} = b_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n \quad (1)$$

Here, $\widehat{Y}$ represents the predicted As concentration; $X_1, X_2, \ldots, X_n$ are the spectral intensities at the selected key wavelengths; and b_0 , $a_1, a_2, \ldots, a_n$ are the regression coefficients. The initial predictions from the TabPFN model are constrained within a reasonable range by introducing boundary factors, namely a lower bound factor (α) and an upper bound factor (β). The lower bound is defined as $\alpha \times \hat{Y}$, and the upper bound as $\beta \times \hat{Y}$, which serve to refine the outputs of the TabPFN model. During the prediction stage, the raw outputs of the TabPFN model are constrained to this interval via a hard boundary operation, whereby predictions falling outside the range are clipped to the corresponding bounds. The effectiveness of the LC module strongly depends on three key hyperparameters: the number of selected wavelengths (n), the lower bound factor (α), and the upper bound factor (β). To identify the optimal hyperparameter combination, we conducted hyperparameter optimization experiments under three spectral views. The detailed optimization process was provided in Supplementary Text S1. The LC-TabPFN model was applied under each spectral view, and its performance was evaluated. For clarity, LC-TabPFN-0, LC-TabPFN-1, and LC-TabPFN-2 refer to the corresponding TabPFN models with linear constraints incorporated under the same three spectral views.

A multi-view linearly constrained TabPFN model (MLC-TabPFN) was developed by integrating the prediction results from the LC-TabPFN models of the original spectral view, first derivative view, and second derivative view. These view models are independently trained, and the fusion is performed after training by computing fixed weights based on the validation-set root mean square error (RMSEv). The fusion weights were assigned according to the root mean square error on the validation set (RMSEv), giving higher weights to perspectives with lower prediction errors. The fusion method follows an inverse error weighting scheme, where the weight for perspective k is defined in Eq. (2).

$$\omega_k = \frac{1/RMSEv_k}{\Sigma_{k=1}^{3} 1/RMSEv_k} \quad (2)$$

Here, k represents the different views (0 = original spectra, 1 = first derivative, 2 = second derivative), and $RMSEv_k$ denotes the validation error of the model trained with perspective k.

3. Results and discussion

3.1. As content and LIBS spectra

P. vittata, a well-known As hyperaccumulator, predominantly accumulates As in its fronds [3]. Therefore, this study focused on determining As concentrations in the fronds. As concentrations were determined by ICP-MS and used as reference values for subsequent LIBS calibration. As shown in Fig. 2(a), concentrations ranged from 9.62 to 1955.74 mg/kg, with an average of 421.9 mg/kg. The wide variation in accumulation levels among different samples provided a useful concentration gradient for LIBS quantification. As shown in Fig. 2(b), three As emission lines were clearly identified at 228.81 nm, 234.98 nm, and 238.12 nm. These emission lines are in close agreement with the excitation lines listed in the National Institute of Standards and Technology (NIST) atomic database, confirming the reliability of the detection results. In addition to As, strong characteristic peaks of Mn and Fe were also observed. This observation is consistent with previous reports [22] indicating that Mn and Fe are essential nutrient elements for P. vittata, and that their uptake by the plant increases in parallel with As accumulation.

3.2. Multivariate models performance

To establish a rapid and accurate model for quantifying As content in *P. vittata*, this study evaluated six multivariate calibration models: TabPFN, RF, XGBoost, CatBoost, PLSR, and ELM. Each original LIBS spectrum contains 1024 features, so each sample's spectral data consists of 1024 features when input into the model. The detailed performance metrics for each model are summarized in Table 1. RF and XGBoost, as ensemble algorithms based on decision trees, exhibited such high flexibility that they almost memorized the calibration samples during calibration, achieving nearly perfect fitting indicators (R^2c=0.988 and 0.999, respectively). However, this excessive learning of noise and sample-specific features from the calibration data resulted in poor generalization, as reflected by the sharp performance drop on the prediction set (R^2p = 0.803 and 0.657; RMSEp=128.999 mg/kg and 170.411 mg/kg; MAPEp=15.642 % and 25.503 %). This makes them unreliable for practical quantitative analysis. CatBoost also performed unsatisfactorily, with a calibration set coefficient of determination (R^2c) of only 0.874. Although its performance improved in the Prediction set (R^2p = 0.929), the prediction error remained high (MAPEp=16.446 %). This may be due to CatBoost's built-in regularization mechanism, which might have been too strict for the present dataset, oversimplifying the model and causing the loss of critical information. In contrast, PLSR demonstrated excellent robustness. By extracting the most relevant latent variables from the high dimensional spectral data, it maintained high accuracy on the prediction set (R^2p = 0.933, MAPEp=11.131 %). ELM's one step learning mechanism, which avoids complex parameter optimization, also achieved comparable results (R^2p = 0.935, MAPEp=11.004 %), performing on par with PLSR.

Notably, TabPFN outperformed all other models. It achieved a prediction set R^2p of 0.956, an RMSEp of 61.340 mg/kg, and the lowest MAPEp of only 9.081 %. As shown in Fig. 3(a), its predicted values are in high agreement with the measured values, with the scatter points closely following the 1:1 reference line. Furthermore, the metric visualization in Fig. 3(b) directly illustrates its superior error control compared to the other models. This exceptional performance stems from TabPFN's novel in-context learning paradigm. As a pre trained Transformer, TabPFN has already learned general patterns from vast synthetic datasets. When applied to our data, it infers the underlying functional relationships by treating the entire dataset as a single prompt, bypassing traditional iterative training. This mechanism allows TabPFN to maintain strong generalization while precisely capturing the essential features of the

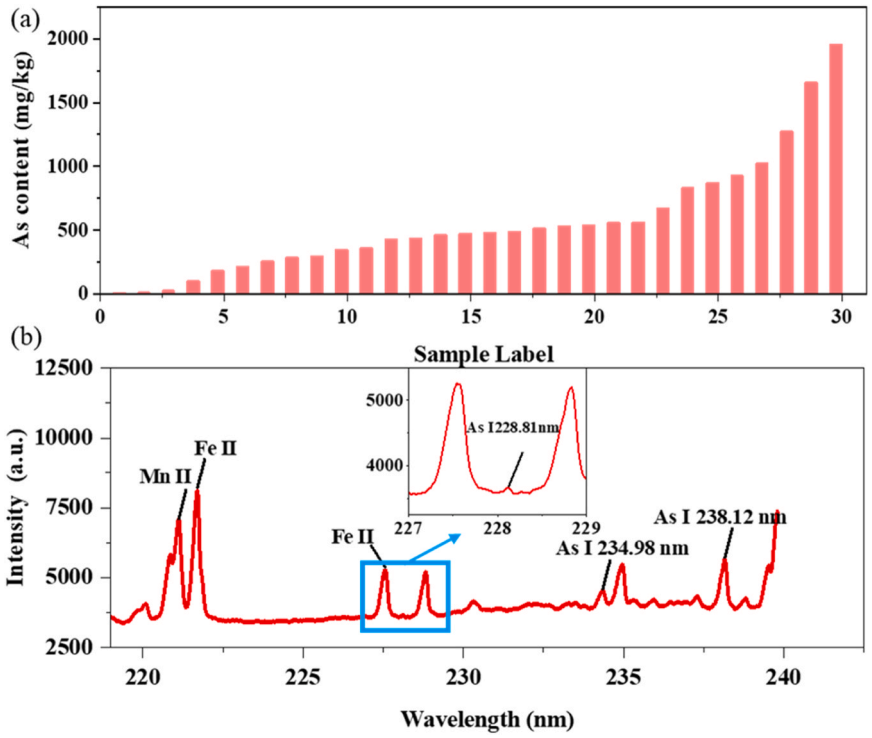


Fig. 2. As concentration and LIBS spectra of *P. vittata*: (a) As content in *P. vittata* fronds. (b) LIBS spectra.

Table 1
Performance of multivariate models for As quantification.

Model	R^2c	RMSEc (mg/kg)	MAPEc (%)	R^2p	RMSEp (mg/kg)	MAPEp (%)
TabPFN	0.999	16.667	2.880	0.956	61.340	9.081
RF	0.988	51.407	23.566	0.803	128.999	15.642
XGBoost	0.999	8.995	8.494	0.657	170.411	25.503
CatBoost	0.874	163.907	228.608	0.929	77.622	16.446
PLSR	0.999	0.006	0.003	0.933	75.503	11.131
ELM	0.999	9.751	14.335	0.935	74.211	11.004

data. Overall, TabPFN comprehensively outperformed all other multivariate models, confirming its considerable potential and advanced capability as a foundation for further optimization.

3.3. MLC-TabPFN model performance

To comprehensively evaluate the performance of the proposed MLC-TabPFN model, we first established a baseline by analyzing the foundational TabPFN model with different spectral views. The predictive performance of TabPFN using the original, first-derivative, and second-derivative spectra as distinct input views is illustrated in Fig. 4(a). Different spectral views provide the TabPFN model with differentiated feature information, which directly affects its learning performance. The original spectra retain the most complete spectral information, which reflects the overall spectral characteristics of As in *P. vittata* fronds. Based on this view, the TabPFN model achieved competitive baseline performance, with prediction set metrics of $R^2p = 0.956$, RMSEp= 61.340 mg/kg, and MAPEp= 9.081 %. However, the presence of inherent interferences such as baseline drift and background noise within the original spectra is a known factor that can limit further improvements in model performance [23]. The first derivative view enhances the characteristic peaks by emphasizing the slope variations in the spectra, which strengthens the trend of the spectral changes and facilitates detecting the peak positions [24]. This is theoretically expected to improve the identification of key spectral lines. However, in this study, the model performance under this view did not surpass that of the original spectra. Specifically, the R^2p dropped to 0.940, while RMSEp increased to 71.478 mg/kg and MAPEp rose to 10.896 %. This indicates that, while the first derivative did enhance some features, it likely amplified noise in certain spectral bands, which negatively impacted the model's overall predictive performance. In contrast, the second derivative view exhibited the best performance. The second derivative spectra effectively eliminate background interference, separate overlapping peaks, and enhance weak signals [25]. This refinement of relevant information allows the model to focus more on spectral features that are highly correlated with As content This focus resulted in the optimal prediction outcome, achieving an R^2p of 0.970, an RMSEp lowered to 50.070 mg/kg, and a MAPEp of 8.537 %. This finding provides strong evidence that the second-derivative view, by filtering out irrelevant data, allows the TabPFN model to build a more accurate quantitative model based on key spectral features of As. However, while transforming the spectral view can reveal more information, the purely data-driven TabPFN model still faces limitations in learning the fundamental relationship between As concentration and spectra from complex data. To overcome this issue and guide the model toward learning more physically interpretable features, we introduced a linearly constrained TabPFN model (LC-TabPFN) based on physical priors.

3.3.1. Contribution of the linear constraint module

Before evaluating the overall performance of the proposed LC-TabPFN model, it was essential to first establish the optimal operating hyperparameters for its core component, the linear constraint (LC) module. The effectiveness of the LC module depends on three key hyperparameters: the number of selected wavelengths (n), the lower bound factor (α), and the upper bound factor (β). To determine the optimal configuration, hyperparameter optimization experiments were carried out under three spectral views. The detailed results are shown in Figures S1–S3, from which it is evident that the best performance across all views was consistently obtained when the hyperparameters were set to $n = 100$, $\alpha = 0.5$, and $\beta = 2.4$.

To clearly demonstrate the effectiveness of the LC module, Table 2

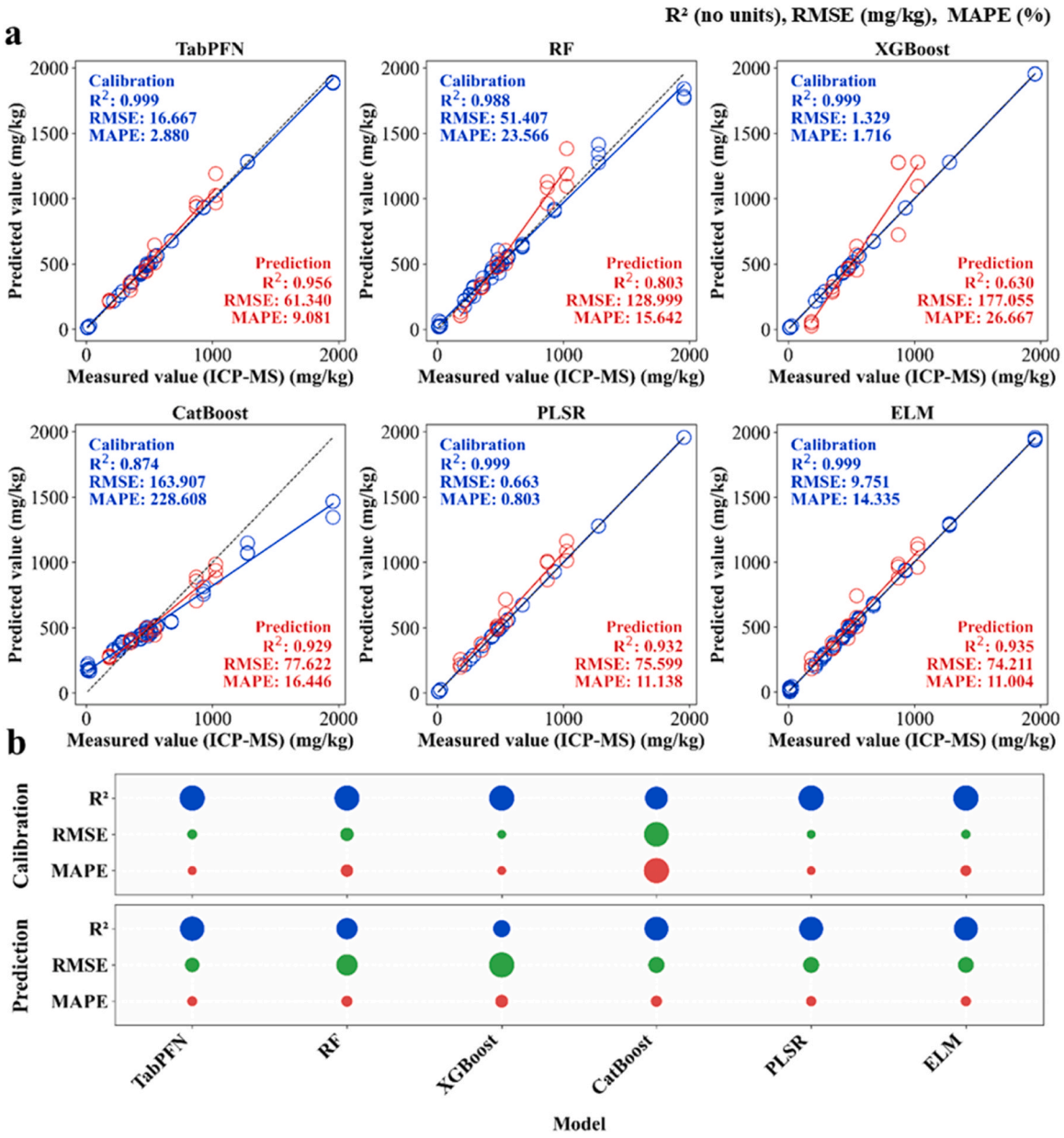


Fig. 3. Performance evaluation of multivariate models. (a) Relationship between measured and predicted As content using different multivariate models. (b) Visualization of R^2, RMSE (mg/kg), and MAPE (%) for calibration and prediction sets.

quantifies the performance gains of LC-TabPFN relative to its reference model (TabPFN without constraints) under three spectral views. On the original and first-derivative spectra, the module reduced prediction errors: RMSEp and MAPEp decreased by 0.302 mg/kg and 0.67 % for the original spectra, and by 0.123 mg/kg and 0.14 % for the first derivative. While these improvements may seem modest, the introduction of the LC module brings significant practical significance. Specifically, it ensures that the model's predictions remain within a physically plausible range by preventing extreme predictions that would be inconsistent with known physical principles. This is particularly important for applications where model reliability and stability in complex biological matrices are crucial, as it helps avoid misdetections caused by black-box models producing predictions that fall outside of physical boundaries. In practical phytoremediation monitoring, even a single extreme prediction could lead to severely misguided decisions regarding harvest timing or remediation efficacy assessment. The LC module effectively mitigates this risk by constraining all predictions within a physically justified range derived from key spectral features. The advantages of the LC module were particularly pronounced in the second-derivative spectra. Here, the module achieved the most substantial reductions in RMSEp and MAPEp, by 0.719 mg/kg and 0.90 % respectively, while also improving R^2p by 0.001. These results confirm that incorporating the LC module consistently improved prediction performance across all spectral views, demonstrating that constraining predictions with physical priors provides a general and effective optimization strategy.

With the optimal hyperparameter setting, LC-TabPFN improved prediction performance across all spectral views, with the best results obtained under the second-derivative view ($R^2p = 0.971$, RMSEp = 49.351 mg/kg, MAPEp = 7.641 %), as shown in Fig. 4(b). By constraining the original TabPFN predictions within a physically reasonable interval, defined relative to a theoretical approximation derived from key wavelengths, the LC module effectively corrected predictions that would otherwise deviate from the expected physical behavior.

3.3.2. *Contribution of the multi-view fusion module*

Although the LC module improved prediction accuracy under each

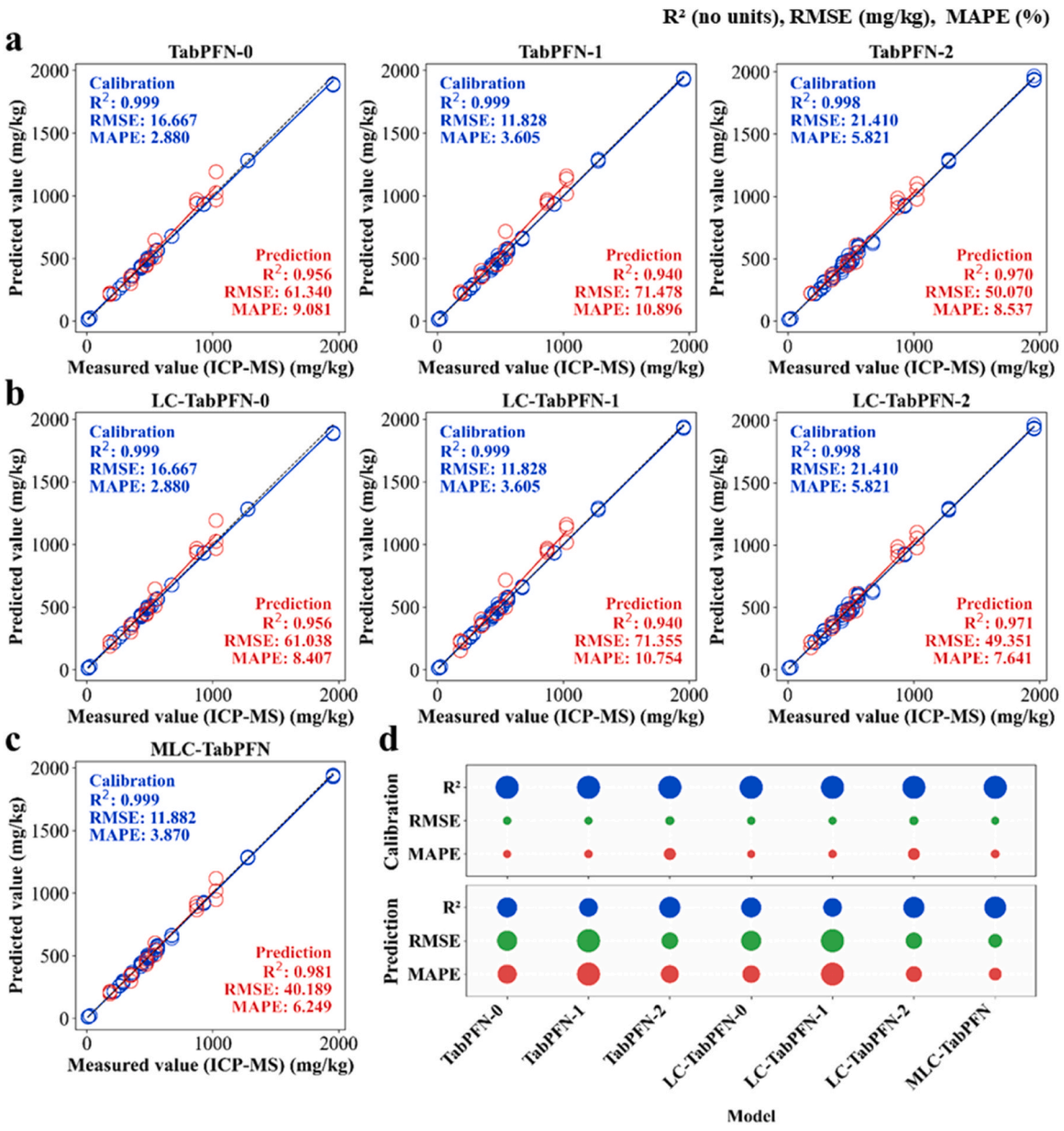


Fig. 4. Performance evaluation of TabPFN, LC-TabPFN, and MLC-TabPFN, where the models are trained under different spectral views: original (0), first derivative (1), and second derivative (2). (a) Predicted versus measured As content for TabPFN models under three spectral views. (b) Predicted versus measured As content for LC-TabPFN models under three spectral views. (c) Predicted versus measured As content for the MLC-TabPFN model. (d) Comparison of R^2, RMSE (mg/kg), and MAPE (%) for calibration and prediction sets across all models.

Table 2
Performance improvements (Δ) of LC-TabPFN and MLC-TabPFN relative to their reference models under different spectral views.

Improved model	Reference model	ΔR^2p	ΔRMSEp (mg/kg),	ΔMAPEp (%)
LC-TabPFN-0	TabPFN-0	0	-0.302	-0.67 %
LC-TabPFN-1	TabPFN-1	0	-0.123	-0.14 %
LC-TabPFN-2	TabPFN-2	0.001	-0.719	-0.90 %
MLC-TabPFN	LC-TabPFN-0	0.025	-20.849	-2.16 %
MLC-TabPFN	LC-TabPFN-1	0.041	-31.166	-4.51 %
MLC-TabPFN	LC-TabPFN-2	0.01	-9.162	-1.39 %

individual spectral view, a single representation still limits the extent of information utilization. To further enhance modeling performance and mitigate the shortcomings of individual spectral view, we developed a multi-view fusion module. Unlike simple aggregation, the fusion approach integrates the complementary information contained in the original, first-derivative, and second-derivative spectra within a unified model, thereby capturing richer and more diverse spectral–concentration relationships. In the multi-view fusion module, the RMSEv for the original spectral, first-derivative, and second-derivative views are 97.248 mg/kg, 110.424 mg/kg, and 132.778 mg/kg, respectively, resulting in fusion weights of 0.383, 0.337, and 0.280. The view with the best performance may receive a lower weight, which could be due to the specificities of the validation and prediction set data. Our fusion algorithm effectively avoids over-reliance on any single preprocessing method.

As summarized in Table 2, the MLC-TabPFN achieved consistent gains over the LC-TabPFN under all three views. For instance, compared with LC-TabPFN under the original and first-derivative spectra view, the fusion model increased R^2p by 0.025 and 0.041, while RMSEp and MAPEp were reduced by 20.849 mg/kg (−2.16 %) and 31.166 mg/kg

(−4.51 %), respectively. Even relative to the best single-view model (LC-TabPFN-2), the fusion model still produced clear improvements. As shown in Fig. 4(c), MLC-TabPFN reached an R^2p of 0.981, with RMSEp and MAPEp further reduced to 40.189 mg/kg and 6.249 %. Fig. 4(d) confirms that multi-view fusion effectively compensates for the deficiencies of individual spectral view and, when combined with linear constraints, delivers the best overall performance in quantitative analysis of As content. The improvements can be explained by the complementary nature of different spectral views: original spectra preserve the full intensity information, first-derivative spectra highlight variation trends and peak positions, and second-derivative spectra suppress background interference. This complementary effect accounts for the superior prediction accuracy of MLC-TabPFN in As quantification. The superior performance of MLC-TabPFN demonstrates that exploiting complementary spectral representations can substantially advance quantitative LIBS analysis, offering a robust framework for trace element determination in complex matrices.

3.4. Discussion

3.4.1. Rationale for selecting TabPFN as the baseline model

TabPFN was selected as the foundational model in this study because of its demonstrated strength in handling small-scale tabular datasets [20], which closely reflects the typical size of LIBS spectral data in practical chemical analysis. In LIBS experiments, the number of samples is often limited to a few hundred or a few thousand spectra due to experimental constraints, making large training sets unrealistic. It should be noted that the small-sample setting of TabPFN refers to small to medium-sized datasets, typically within 10,000 samples. Owing to the high cost and complexity of acquiring reliable chemical reference values (e.g., ICP-MS measurements), LIBS datasets in environmental and plant analysis rarely exceed this scale. Therefore, the data volume in typical LIBS applications falls well within the intended operating range of TabPFN. In practical LIBS applications, models need to be deployed quickly. TabPFN meets this requirement because its inference is fast, relying only on a direct forward pass once pretrained, and it does not require complex feature engineering [20]. Unlike many deep learning models that require extensive hyperparameter tuning and large training sets, TabPFN has been shown to achieve competitive performance without complex optimization [20], making it highly suitable for practical chemical analysis scenarios. In our work, TabPFN provided a strong baseline across different spectral views, reaching $R^2p = 0.956$, RMSEp= 61.340 mg/kg, and MAPEp= 9.081 % when applied to the original spectra. These results confirm that even in the presence of baseline drift and spectral noise, TabPFN can effectively capture the key relationships between spectral features and As concentration. These findings highlight TabPFN's suitability as a backbone model for LIBS-based quantitative analysis, providing both robustness under small-sample conditions and flexibility for integration with domain-specific enhancements.

3.4.2. Analysis of the core components in MLC-TabPFN

The introduction of the linear constraint (LC) module in TabPFN is more than a minor adjustment of prediction outcomes. It embeds the prior knowledge that spectral intensity and concentration are positively correlated into an otherwise purely data driven framework. The purpose of this physical prior is not to fully explain the LIBS physical process, but rather to constrain the model's predictions, ensuring they do not deviate from basic physical principles. Although the TabPFN model is effective in learning complex patterns from spectral data, its black-box nature ignores spectroscopic principles, which can lead to spurious correlations learned from noise or physically implausible predictions when applied to samples outside the training distribution. Similar challenges have been reported in LIBS quantitative analysis. Song et al. proposed the dominant factor kernel extreme learning machine (DF-KELM), which improves reliability through domain knowledge based spectral line modeling and nonlinear residual correction; for example, the R^2 of carbon prediction increased from 0.861 (PLSR) to 0.921, and RMSEp decreased from 3.15 % to 2.46 % [26]. They later developed the spectral knowledge-based regression (SKR) algorithm, which integrates knowledge guided linear modeling with kernel based nonlinear regression and achieved better performance than baseline methods across multiple datasets [12]. These studies indicate that combining data driven and knowledge driven approaches can simultaneously enhance prediction accuracy and interpretability. The LC module follows the similar principle by introducing the linear relationship between key As related wavelengths and concentration to restrict the prediction range and prevent departures from reality. As shown in Table 2, the LC module consistently improves performance across three spectral views. The improvement is most evident for the second derivative spectra, where RMSEp was reduced by 0.719 mg/kg and MAPEp by 0.90 %. This demonstrates that even in a highly optimized base model, constraining predictions with physical priors can still provide additional accuracy and optimization space.

The advantages of MLC-TabPFN over purely "black-box" methods lie in its ability to incorporate scientific rigor, interpretability, and robustness. By embedding physical priors, such as the positive correlation between spectral intensity and concentration, MLC-TabPFN avoids the spurious correlations that black-box models might learn, ensuring that predictions align with physical principles rather than just minimizing mathematical error. Unlike traditional black-box models, our approach allows for clear explanations of predictions, such as why they fall within certain ranges determined by physical constraints. This interpretability is crucial in practical applications, particularly to avoid inappropriate soil remediation actions based on inaccurate predictions. Furthermore, the physical constraints introduced by the LC module help prevent extreme predictions, enhancing the robustness of the model, especially when facing anomalous spectra, and preventing overfitting by maintaining predictions within physically plausible bounds.

The introduction of the multi-view fusion module is motivated by the fact that different spectral views provide complementary rather than redundant information. Original spectra retain overall intensity information, while first derivatives emphasize slope and peak features and second derivatives reduce baseline drift, separate overlapping peaks, and highlight weak signals. Together these views provide complementary information that enables a more complete relationship between spectra and concentration. Previous studies have provided direct evidence for this complementarity. Zhang et al. [14] proposed a stacking preprocessing ridge regression (SPRR) method for near-infrared spectral analysis, in which original spectra, first- and second-derivative spectra, SNV, and MSC were combined. Their correlation analysis showed that these preprocessing results contained significant complementary information. As a result, SPRR outperformed single preprocessing and existing ensemble methods across six datasets (apple, meat, wheat, olive oil, tablets, and maize), achieving lower RMSE and higher R^2. This outcome supports the view that multi-view integration can effectively exploit complementary information to improve model performance. Our results are consistent with this observation. As shown in Table 2, compared with the best single-view model, multi-view fusion further improved prediction accuracy, achieving $R^2p = 0.981$, RMSEp= 40.189 mg/kg, and MAPEp= 6.249 %. These results demonstrate that the fusion module maximizes the utility of different spectral views and compensates for the limitations of any single representation. Multi-view fusion thus provides an effective ensemble learning strategy for spectroscopic analysis and improves the accuracy of As quantification.

Although the MLC-TabPFN model showed excellent performance on the calibration set (R^2c=0.999), the slight decrease in performance on the validation set (R^2v=0.959) and prediction set ($R^2p = 0.981$) suggests a potential risk of overfitting. However, TabPFN's in-context learning approach, which does not rely on backpropagation, helps mitigate this risk compared to traditional deep learning models.

Additionally, the LC module, by introducing physical prior knowledge, ensures that the model's predictions stay within a physically plausible range, further reducing overfitting. The model's performance on the prediction set ($R^2p = 0.981$, RMSEp=40.189 mg/kg, and MAPEp=6.249 %) demonstrates that even on an entirely independent prediction set, the model maintains high precision and effectively controls overfitting risk, without compromising its generalization ability.

3.4.3. Limitations and future outlook

Although the proposed MLC-TabPFN model showed excellent performance in the quantitative analysis of As in *P. vittata*, several limitations remain and should be addressed in future studies. First, the modeling and validation in this work were restricted to As in *P. vittata*. While the model achieved promising results on this dataset, its generalization ability has not been fully evaluated. Whether it can be applied to other plant species (e.g., rice, vegetables), different environmental matrices (e.g., soil, water sediments), and other heavy metals (e.g., cadmium, lead, mercury) requires systematic assessment using broader datasets. Future research should therefore focus on collecting more diverse samples to examine the adaptability and robustness of the model across species, matrices, and elements. Furthermore, future research could also include plants at different growth stages to evaluate how elemental concentrations change during plant growth, and increase the number of plants and replicates to improve sample diversity and representativeness. This would address the current limitation of sample size and enhance the model's stability and generalization ability. Second, although LIBS has strong potential for in-situ analysis, this study employed pelletized samples to obtain a flat, homogeneous, and reproducible measurement surface, thereby reducing spectral fluctuations caused by the structural heterogeneity of plant tissues. As a result, the current validation was performed under controlled laboratory conditions rather than true in-situ scenarios. Future studies will investigate extending the proposed MLC-TabPFN framework to in-situ LIBS measurements with minimal or no sample preparation, aiming to enhance its practical applicability. Third, the LC module was constructed on a simplified physical approximation. Although the module improved the physical plausibility and accuracy of predictions in our experiments, its theoretical basis remains limited. Future work could explore the use of more complex and precise physical models, such as plasma emission-based theoretical formulations or corrections that account for self-absorption effects, to achieve deeper integration of physical prior knowledge with machine learning and further improve model interpretability and predictive performance.

Despite these limitations, the central contribution of this study is the development of the MLC-TabPFN model, which combines data-driven learning with knowledge-driven. Through systematic experiments, we demonstrated that the model integrates complementary information from multiple spectral views and constrains predictions with linear physical priors, achieving performance in LIBS-based As quantification of *P. vittata* that is superior to both the baseline TabPFN and several mainstream algorithms. This work provides an effective solution for the rapid and accurate determination of trace elements in complex biological matrices and highlights the potential of combining physical knowledge with advanced machine learning in spectroscopic quantitative analysis.

4. Conclusions

This study proposes a novel quantitative analysis model, MLC-TabPFN, which combines data-driven and knowledge-driven approaches for the rapid and accurate determination of As content in *P. vittata* fronds using LIBS. Initially, we systematically evaluated several multivariate models, including RF, XGBoost, CatBoost, PLSR, ELM, and TabPFN. Among them, TabPFN, based on the in-context learning paradigm, delivered the best baseline performance on small sample LIBS datasets, achieving $R^2p = 0.956$, RMSEp= 61.340 mg/kg, and MAPEp= 9.081 %, outperforming conventional methods. Building on this foundation, we proposed the MLC-TabPFN framework with two main innovations. First, the linear constraint module introduced the physical prior that spectral intensity should be positively correlated with concentration, thereby correcting prediction behavior and improving both plausibility and accuracy. In the optimal second-derivative spectral view, this module reduced RMSEp by 0.719 mg/kg and decreased MAPEp by 0.90 %, while maintaining a high R^2p of 0.971. Second, the multi-view fusion module integrated complementary information from original, first-derivative, and second-derivative spectra, overcoming the limitations of single views. Compared with the best single-view model, fusion improved R^2p from 0.971 to 0.981 and reduced MAPEp from 7.641 % to 6.249 %.The results confirmed that the MLC-TabPFN model achieved superior accuracy over all other models, reaching a performance level that is competitive with or exceeds the state of the art in this field.

In conclusion, this study successfully developed a novel LIBS quantitative analysis method that merges data-driven learning with physical constraints. By effectively utilizing multi-view spectral information, the MLC-TabPFN model provides a reliable solution for the efficient and precise determination of trace elements in complex biological matrices. This work also demonstrates the broad potential of integrating advanced machine learning with spectroscopic techniques for applications in environmental monitoring and phytoremediation.

Environmental implications

Arsenic contamination constitutes a severe threat to ecosystem security and human health. Phytoremediation with *P. vittata* offers a sustainable solution, yet its widespread application has been hampered by the inability to monitor extraction efficiency effectively. Our MLC-TabPFN model enables rapid, high-precision arsenic quantification within the plant. This critical analytics breakthrough allows for precise optimization of harvest schedules to maximize removal and prevent arsenic re-release. Consequently, it facilitates reliable soil cleanup, accelerates restoration, and mitigates long-term leaching risks. By providing a robust verification tool, our work promotes large-scale phytoremediation deployment, strengthening environmental monitoring and significantly reducing the risks of arsenic pollution.

CRediT authorship contribution statement

Fei Liu: Writing – review & editing, Project administration, Methodology, Conceptualization. **Dongxing Guan:** Supervision, Resources, Data curation. **Zhihua Dai:** Resources, Data curation. **Jing Huang:** Writing – review & editing, Conceptualization. **Lei Feng:** Writing – review & editing, Methodology. **Liuye Cao:** Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Research Project of Zhejiang Key Laboratory of Agricultural Remote Sensing and Information Technology.

Data availability

The authors do not have permission to share data.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2026.141490.

Data availability

The authors do not have permission to share data.

References

[1] Oremland, R.S., Stolz, J.F., 2003. The ecology of arsenic. Science 300, 939–944. https://doi.org/10.1126/science.1081903.

[2] Hettick, B.E., Cañas-Carrell, J.E., French, A.D., Klein, D.M., 2015. Arsenic: a review of the element's toxicity, plant interactions, and potential remediation methods. J Agric Food Chem 63, 7097–7107. https://doi.org/10.1021/acs.jafc.5b02487.

[3] Ma, L.Q., Komar, K.M., Tu, C., Zhang, W., Cai, Y., Kennelley, E.D., 2001. A fern that hyperaccumulates arsenic. Nature 409, 579–5719. https://doi.org/10.1038/35054664.

[4] Zeng, P., Guo, Z., Xiao, X., Peng, C., Feng, W., Xin, L., Xu, Z., 2019. Phytoextraction potential of Pteris vittata L. co-planted with woody species for As, Cd, Pb and Zn in contaminated soil. Sci Total Environ 650, 594–603. https://doi.org/10.1016/j.scitotenv.2018.09.055.

[5] Frank, J., Krachler, M., Shotyk, W., 2005. Direct determination of arsenic in acid digests of plant and peat samples using HG-AAS and ICP-SF-MS. Anal Chim Acta 530, 307–316. https://doi.org/10.1016/j.aca.2004.09.077.

[6] Rajaković, L.V., Marković, D.D., Rajaković-Ognjanović, V.N., Antanasijević, D.Z., 2012. Review: The approaches for estimation of limit of detection for ICP-MS trace analysis of arsenic. Talanta 102, 79–87. https://doi.org/10.1016/j.talanta.2012.08.016.

[7] Pandey, S., Mishra, S., 2025. A review of sensing technologies for arsenic detection in drinking water. Int J Environ Sci Technol 22, 2809–2832. https://doi.org/10.1007/s13762-024-05912-1.

[8] Peng, J., Ye, L., Liu, Y., Zhou, F., Xu, L., Zhu, F., Huang, J., Liu, F., 2024. Characterization of the distribution of mineral elements in chromium-stressed rice (Oryza sativa L.) leaves based on laser-induced breakdown spectroscopy and data augmentation. Spectrochim Acta B Spectrosc 222, 107072. https://doi.org/10.1016/j.sab.2024.107072.

[9] Wang, W., Man, Z., Li, X., Chen, R., You, Z., Pan, T., Dai, X., Xiao, H., Liu, F., 2023. Response mechanism and rapid detection of phenotypic information in rice root under heavy metal stress. J Hazard Mater 449, 131010. https://doi.org/10.1016/j.jhazmat.2023.131010.

[10] Li, X., Huang, J., Chen, R., You, Z., Peng, J., Shi, Q., Li, G., Liu, F., 2023. Chromium in soil detection using adaptive weighted normalization and linear weighted network framework for LIBS matrix effect reduction. J Hazard Mater 448, 130885. https://doi.org/10.1016/j.jhazmat.2023.130885.

[11] Xie, W., Xu, J., Huang, L., Xu, Y., Wan, Q., Chen, Y., Yao, M., 2024. Improved Cd detection in rice grain using LIBS with husk-based XGBoost transfer learning. Agriculture 14, 2053. https://doi.org/10.3390/agriculture14112053.

[12] Song, W., Afgan, M.S., Yun, Y.-H., Wang, H., Cui, J., Gu, W., Hou, Z., Wang, Z., 2022. Spectral knowledge-based regression for laser-induced breakdown spectroscopy quantitative analysis. Expert Syst Appl 205, 117756. https://doi.org/10.1016/j.eswa.2022.117756.

[13] Mishra, P., Biancolillo, A., Roger, J.M., Marini, F., Rutledge, D.N., 2020. New data preprocessing trends based on ensemble of multiple preprocessing techniques. TrAC Trends Anal Chem 132, 116045. https://doi.org/10.1016/j.trac.2020.116045.

[14] Huang, H., Fang, Z., Xu, Y., Lu, G., Feng, C., Zeng, M., Tian, J., Ping, Y., Han, Z., Zhao, Z., 2024. Stacking and ridge regression-based spectral ensemble preprocessing method and its application in near-infrared spectral analysis. Talanta 276, 126242. https://doi.org/10.1016/j.talanta.2024.126242.

[15] Wang, W., Man, Z., Li, X., Zhao, Y., Chen, R., Pan, T., Wang, L., Dai, X., Xiao, H., Liu, F., 2024. Multi-phenotype response and cadmium detection of rice stem under toxic cadmium exposure. Sci Total Environ 917, 170585. https://doi.org/10.1016/.scitotenv.2024.170585.

[16] Zhou, J., Huang, W., Sattar, H., Hu, Q., Liu, H., Wang, G., Zhang, D., Guo, L., 2026. LIBS matrix effect deviation compensation through acoustic-optical spectra fusion LIBS technique. Talanta 296, 128453. https://doi.org/10.1016/j.talanta.2025.128453.

[17] Kabir, M.H., Guindo, M.L., Chen, R., Liu, F., 2021. Geographic origin discrimination of millet using Vis-NIR spectroscopy combined with machine learning techniques. Foods 10, 2767. https://doi.org/10.3390/foods10112767.

[18] Rehan, I., Rehman, M.U., Aamir, M., Islam, S., 2025. A CatBoost and ExtraTrees-based softvoting ensemble approach for non-invasive diabetes detection using hair LIBS spectral data. Microchem J 217, 114980. https://doi.org/10.1016/j.microc.2025.114980.

[19] Wang, W., Kong, W., Shen, T., Man, Z., Zhu, W., He, Y., Liu, F., 2021. Quantitative analysis of cadmium in rice roots based on LIBS and chemometrics methods. Environ Sci Eur 33, 37. https://doi.org/10.1186/s12302-021-00480-4.

[20] Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S.B., Schirrmeister, R.T., Hutter, F., 2025. Accurate predictions on small data with a tabular foundation model. Nature 637, 319–326. https://doi.org/10.1038/s41586-024-08328-6.

[21] Messaoud Aberkane, S., Safi, A., Botto, A., Campanella, B., Legnaioli, S., Poggialini, F., Raneri, S., Rezaei, F., Palleschi, V., 2020. Laser-induced breakdown spectroscopy for determination of spectral fundamental parameters. Appl Sci 10, 4973. https://doi.org/10.3390/app10144973.

[22] Cao, X., Ma, L.Q., Tu, C., 2004. Antioxidative responses to arsenic in the arsenic-hyperaccumulator Chinese brake fern (Pteris vittata L.). Environ Pollut 128, 317–325. https://doi.org/10.1016/j.envpol.2003.09.018.

[23] Xia, Z., Che, X., Ye, L., Zhao, N., Guo, D., Peng, Y., Lin, Y., Liu, X., 2023. A synergetic strategy for brand characterization of colla corii asini (Ejiao) by LIBS and NIR combined with partial least squares discriminant analysis. Molecules 28, 1778. https://doi.org/10.3390/molecules28041778.

[24] Yu, L., Li, J., Wang, J., Li, J., Li, J., Xi, Q., Han, Z., 2024. Derivative spectroscopy and its application at detecting the weak emission/absorption lines. arXiv Prepr arXiv 2404, 03866. https://doi.org/10.48550/ARXIV.2404.03866.

[25] Czarnecki, M.A., 2015. Resolution enhancement in second-derivative spectra. Appl Spectrosc 69, 67–74. https://doi.org/10.1366/14-07568.

[26] Song, W., Hou, Z., Gu, W., Afgan, M.S., Cui, J., Wang, H., Wang, Y., Wang, Z., 2022. Incorporating domain knowledge into machine learning for laser-induced breakdown spectroscopy quantification. Spectrochim Acta B Spectrosc 195, 106490. https://doi.org/10.1016/j.sab.2022.106490.