Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Geographical traceability of intact Astragali Radix slices from six producing regions by data fusion of laser-induced breakdown spectroscopy and Raman spectroscopy

Yikang Hou
China People's Public Security University
Zhuoxi Li
China People's Public Security University
Jie Lian
rlianjie@163.com

China People's Public Security University

Research Article

Keywords: Astragali Radix, laser-induced breakdown spectroscopy, Raman spectroscopy, data fusion, geographical traceability, in situ analysis

Posted Date: May 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9261462/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Geographical origin is a major determinant of Astragali Radix quality, yet rapid origin trac- ing of commercial slices remains difficult because most established methods require destruc- tive pretreatment. Here, 180 intact Astragali Radix slices from six producing regions were analyzed by paired laser-induced breakdown spectroscopy (LIBS) and Raman spectroscopy directly on the cross section, without grinding or pelletizing. Measurements were collected from the pith, xylem ring, and cortex, yielding 8100 LIBS spectra and 1620 Raman spectra. To integrate elemental and molecular information, support vector machine (SVM), random forest (RF), and dual-stream one-dimensional convolutional neural network (1D-CNN) mod- els were developed and evaluated over five independent sample-level random splits. The fused SVM model gave a mean external-test accuracy of 94.67% (SD 1.83%), whereas the fused dual-stream 1D-CNN achieved mean cross-validation and external-test accuracies of 99.87% (SD 0.30%) and 99.33% (SD 1.49%), respectively. In the deep-learning framework, all 8100 single-pulse LIBS spectra were retained rather than averaged, allowing local spatial heterogeneity of intact slices to be exploited. Grad-CAM indicated that Ca, Mg, Na, and K emission lines together with Raman bands at 1118, 1266, 1336, and 1457 cm__1 associ- ated with cellulose and lignin made the strongest contributions to discrimination. These results demonstrate that LIBS–Raman data fusion enables rapid and accurate geographical traceability of Astragali Radix slices under minimal pretreatment while preserving a chemi- cally interpretable basis for classification. Further validation across harvest years and larger cohorts is still required.

1. Introduction

Astragali Radix, the dried root of *Astragalus membranaceus* or *A. membranaceus* var. *mongholicus*, is one of the most widely used tonic herbal medicines in East Asia and an im-portant medicinal–edible resource. Its pharmacological value is closely related to flavonoids, saponins, polysaccharides, and inorganic elements, which collectively contribute to immunomod- ulatory, anti-inflammatory, antioxidant, and cardiovascular-protective activities [1, 2, 3]. At the same time, the quality of Astragali Radix is strongly origin-dependent. Soil mineral com- position, hydrothermal regime, altitude, illumination, and rhizosphere environment jointly

influence elemental uptake and metabolic regulation in the root, ultimately generating region- specific chemical phenotypes [4, 5, 7]. As the commercial circulation of Astragali Radix expands, origin mislabeling and source ambiguity have become persistent quality-control problems. A rapid and reliable provenance method that operates on commercial slices rather than laboratory-prepared powders is therefore of clear practical importance.

Current origin-authentication strategies for Astragali Radix mainly rely on DNA barcod- ing, chromatography, NMR, and elemental analysis [8, 9, 6, 5]. These methods provide strong genetic or chemical evidence, but they usually require grinding, extraction, digestion, or other labor-intensive pretreatment steps. Their throughput and field compatibility are therefore limited, especially for routine

screening of decoction slices [8, 9, 10]. This limitation has inten- sified interest in spectroscopic approaches capable of capturing integral chemical fingerprints under minimal pretreatment.

Among rapid spectroscopic tools, Raman spectroscopy and LIBS represent two comple- mentary analytical routes. Raman spectroscopy probes molecular vibrations and is sensitive to cellulose, lignin, and related organic structures; it has consequently been applied to herbal quality evaluation and origin-related classification [11, 10]. Its performance in complex plant matrices, however, can be compromised by fluorescence background and weak-band overlap. LIBS, in contrast, directly records elemental emission from laser-induced microplasmas and has shown considerable promise for origin authentication of agricultural and herbal materials because it rapidly captures multielement information with little or no sample preparation [12, 13, 14]. For Astragali Radix, whose provenance is shaped by both mineral nutrition and organic structural composition, either modality alone is likely to provide only a partial description of origin-related differences.

Multisource spectral fusion offers a direct way to address this limitation. Previous stud- ies have shown that when different spectroscopic modalities encode complementary chemical levels, fused models usually outperform single-modality models [15, 16, 17]. In the context of Astragali Radix, LIBS–IR fusion has already demonstrated that elemental and molecular in- formation are mutually informative for regional discrimination [12], and recent LIBS–Raman studies further confirmed the analytical value of this strategy [18, 19]. Even so, two is- sues remain insufficiently resolved. First, most available studies rely on powders, pellets, or relatively small market-derived sample sets, so the role of tissue differentiation and spatial heterogeneity in intact commercial slices remains unclear. Second, improvements in classifica- tion accuracy are often reported without a sufficiently explicit discussion of the spectroscopic basis of discrimination.

The present study addresses both issues. A total of 180 intact Astragali Radix slices from six producing regions were analyzed directly on the cross section by paired LIBS and Raman measurements. Traditional machine-learning models and a dual-stream 1D-CNN were developed under strict sample-level partitioning, and Grad-CAM was used to trace the learned decisions back to specific elemental lines and Raman bands. The aims were:

(i) to establish a practical in-situ traceability framework for intact Astragali Radix slices under minimal pretreatment; (ii) to determine whether the spatial heterogeneity preserved in intact slices can be transformed into useful discriminative information; and (iii) to clarify the elemental and molecular spectroscopic basis underlying regional discrimination.

2. Materials and methods

2. 1. Samples

A total of 180 authenticated Astragali Radix slices were collected from six producing regions in China: Shaanxi, Heilongjiang, Jilin, Inner Mongolia, Shanxi, and Gansu, with 30 samples from each region. All

samples were kept in their original slice form and were analyzed directly, without grinding, pelletizing, or chemical extraction. Before measurement, the samples were stored in a dry and dark environment, equilibrated to room temperature, and gently cleaned to remove superficial debris.

The cross section of Astragali Radix shows a distinct internal organization. To preserve this structural information while ensuring spatial correspondence between the two modalities, both LIBS and Raman spectra were acquired from three representative anatomical regions: pith, xylem ring, and cortex. Compared with powder-based protocols, this design more closely resembles real inspection conditions for decoction slices while retaining the intrinsic chemical heterogeneity associated with tissue differentiation.

2.2. LIBS and Raman measurements

The sample layout and instrumental setups are shown in Fig. 1. LIBS measurements were performed with a Nd:YAG pulsed laser at 1064 nm using a pulse energy of 96 mJ, a repetition rate of 5 Hz, an integration time of 2 ms, and a gate delay of 1.28 ţs. The delay time was chosen within the range commonly used for plant matrices and was finalized after preliminary optimization based on the signal-to-noise ratios of the main elemental lines [20, 21]. Spectra were collected by a six-channel spectrometer. For each sample, three points were selected in each anatomical region, and five consecutive single-pulse LIBS spectra were acquired at each point while the translation stage shifted the sampling position between pulses. Each sample therefore yielded 45 LIBS spectra, giving 8100 LIBS spectra in total.

Raman spectra were collected with a Renishaw inVia spectrometer using 1064 nm excita- tion, a nominal spectral range of _208.384 to 2580.97 cm__1, a spectral resolution of 1.0 cm__1 , and an integration time of 5000 ms. Only the 300–1800 cm__1 region was used for modeling. The 1064 nm laser was selected primarily to suppress the strong natural fluorescence typical of plant matrices [10]. Raman measurements followed the same region-based sampling layout as LIBS. Five repeated acquisitions were collected at each sampling point and averaged to form the final Raman spectrum for that point. Each sample produced 9 Raman spectra, giving 1620 Raman spectra in total. Accordingly, the two modalities were paired one-to-one at both the sample level and the measurement-point level.

2.3. Spectral pretreatment, data partitioning, and model development

Before modeling, LIBS spectra were processed by Savitzky–Golay smoothing, asymmet- ric least-squares baseline correction, and standard normal variate normalization, followed by retention of the 250–800 nm region [22, 23]. Raman spectra were processed by Savitzky– Golay smoothing, adaptive iteratively reweighted penalized least-squares baseline correction, min–max normalization, and retention of the 300–1800 cm__1 region [22, 24]. Data parti- tioning was performed strictly at the physical-sample level. Thirty samples (five per region)

were randomly assigned to the external test set, and the remaining 150 samples constituted the model-development set. Within the development set, stratified grouped five-fold cross- validation was used, with grouping by sample identity so that different measurement points from the same physical sample never appeared in both training and validation subsets.

Traditional models were built with SVM and RF classifiers [25, 26]. For these models, the five LIBS single-pulse spectra collected at a given point were averaged first and then paired with the corresponding averaged Raman spectrum, giving nine point-level paired observations per sample. Dimensionality reduction and model hyperparameters were optimized within the cross-validation framework; the final parameter settings are listed in Supplementary Table S1.

Deep learning was implemented with a dual-stream 1D-CNN based on residual blocks [27]. LIBS and Raman spectra were fed into two independent convolutional branches with the same topology, and the high-level features were concatenated for final classification. Detailed architectural settings are given in Supplementary Table S2. Unlike the traditional models, the deep model did not average the five LIBS single-pulse spectra collected at each point. Instead, every single-pulse LIBS spectrum was retained as an individual input, and the corresponding Raman spectrum was duplicated five times to preserve the point-level pairing. This yielded 8100 paired LIBS–Raman inputs and allowed local spectral heterogeneity in intact slices to be explicitly used during feature learning. Network training used the AdamW optimizer with an initial learning rate of 3 × 10__4, a weight decay of 1 × 10__4, a batch size of 32, 35 epochs, and ReduceLROnPlateau scheduling [28].

2.4. Data fusion and model evaluation

For SVM and RF, feature-level fusion was implemented by concatenating the reduced LIBS and Raman feature vectors. For the dual-stream 1D-CNN, fusion was performed at the high-feature level after modality-specific feature extraction. Single-modality LIBS and Raman models were also constructed for comparison.

Model performance was assessed by mean (SD) of grouped five-fold cross-validation ac- curacy and by accuracy, precision, recall, and F1 score on the independent external test set. To evaluate partition sensitivity, the entire sample-level split procedure was repeated with five independent random seeds (0, 1, 2, 3, and 42). The robustness of the fused deep model was further assessed by adding 5% Gaussian white noise to the external-test spectra and repeating the perturbation 20 times. The penultimate-layer features were visualized in two dimensions for qualitative inspection, and Grad-CAM was used to identify the spectral variables contributing most strongly to the final classification decisions [29].

3. Results and discussion

3. 1. Spectral characteristics and within-sample heterogeneity

Mean spectra for the six producing regions and the corresponding variance spectra are shown in Fig. 2, and the main assignments are summarized in Table 1. In the LIBS domain, several emission lines associated with Mg, Ca, Na, and K displayed stable intensity differences across regions, indicating that the roots retained measurable signatures of region-specific mineral environments. In the Raman domain, the dominant information arose from vibrations associated with lignin, cellulose, and related oxygen-containing frameworks. Bands near 491, 1118, 1336, 1374, 1460, and 1605 cm__1 were particularly reproducible and consistent with previous vibrational assignments for lignocellulosic materials [30, 31].

The variance spectra further show that the strongest fluctuations coincide with major spectral features rather than with baseline regions, suggesting that a substantial part of the observed variability reflects genuine microscale heterogeneity within intact slices rather than random instrumental noise alone. This interpretation is especially plausible for LIBS, which is intrinsically sensitive to local compositional differences at the ablation spot. In other words, the intact-slice design preserved not only the average chemical signature of each sample but also the tissue-dependent spatial variability embedded in that signature. This observation provides an experimental basis for the single-pulse strategy adopted for the deep model.

Table 1: Main LIBS emission lines and Raman bands observed for *Astragali Radix*

(a)MainLIBSemissionlinesobservedfor*Astragali Radix*

Wavelength (nm)	Assignment	Wavelength (nm)	Assignment
247.83	C I	486.14	H I
279.55	Mg II	588.83	Na I
285.21	Mg I	656.27	H I
309.29	Al I	670.70	Li I
317.57	Ca II	742.36	N I
358.20	Fe I	746.81	N I
388.30	CN	766.45	K I
393.36	Ca II	769.90	K I
396.85	Ca II	777.20	O I
422.69	Ca I		
460.70	Sr I		

(b)Ramanbandsandtentativeassignmentsobservedfor*Astragali Radix*

Raman	Shift (cm−1)	Tentative Assignment
	491	Aromatic ring deformation in lignin [30]
	543	Tryptophan side-chain C–C–C vibration
	640	Guaiacyl ring breathing in lignin [30]
	730	C–S stretching of sulfur-containing groups
	943	C–OH terminal bending in lignin [31]
	1089	Glycosidic C–C–O stretching in cellulose [31]
	1118	Ring-framework C–C stretching in cellulose [31]
	1268	Aryl–O stretching and G-ring vibration in lignin [30]
	1336	HCC/HCO bending in cellulose [31]
	1374	Symmetric CH3 bending in lignin [30]
	1460	CH2 scissoring vibration [30, 31]
	1605	Aromatic C=C stretching in lignin [30]

3.2. Conventional models and the gain from fusion

The representative external-test results for the traditional models (seed = 42) are listed in Table 2. Under single-modality conditions, LIBS generally outperformed Raman, indicat- ing that provenance differences among the six regions are expressed more strongly in the elemental domain than in the molecular domain when shallow models are used. Once the two modalities were fused, however, both SVM and RF improved across all reported met- rics. Across the five independent sample-level random splits, the mean external-test accuracy reached 94.67% (SD 1.83%) for the fused SVM model and 92.00% (SD 1.83%) for the fused RF model (Table 3). These results confirm that Raman-derived molecular information provides a stable complement to the elemental information captured by LIBS.

In the representative split, the fused SVM model misclassified only one of the 30 external- test samples, with a sample from Jilin predicted as Heilongjiang. This error is not surprising from a chemometric perspective because the two northeastern regions share similar climatic backgrounds and soil characteristics. It is also worth noting that the traditional models relied on point-level averaging of the five single-pulse LIBS spectra, which improves stability but inevitably suppresses part of the fine-scale local variation retained in the intact-slice measurements.

Table 2: External-test performance of conventional models in the representative split (seed = 42)

Model	Modality	Accuracy	Precision	Recall	F1score
RF	LIBS	90.00%	92.46%	90.00%	90.15%
RF	Raman	86.67%	89.13%	86.67%	86.36%
RF	Fusion	93.33%	95.24%	93.33%	93.06%
SVM	LIBS	93.33%	94.44%	93.33%	93.27%
SVM	Raman	90.00%	91.11%	90.00%	89.93%
SVM	Fusion	96.67%	97.22%	96.67%	96.63%

3.3. Dual-stream 1D- CNN performance and multiscale information use

Table 3(a) summarizes the grouped five-fold cross-validation results of the dual-stream 1D- CNN in the representative split (seed = 42). The single-modality LIBS and Raman networks reached mean accuracies of 87.33% (SD 10.38%) and 86.00% (SD 6.41%), respectively, with marked fold-to-fold fluctuations. In contrast, the fused dual-stream network achieved 100.00% accuracy in every fold of the representative split, indicating that combined elemental and molecular inputs substantially stabilized the feature-learning process.

The same conclusion holds across repeated sample-level splits. As summarized in Ta- ble 3(b), the fused dual-stream 1D-CNN yielded mean cross-validation and external-test accuracies of 99.87% (SD 0.30%) and 99.33% (SD 1.49%), respectively, with four of the five external-test splits classified perfectly. The superiority of the deep fused model over the tradi- tional fused models was therefore not the product of a favorable partition, but a reproducible pattern across independent randomizations.

Table 3: Performance of the dual-stream 1D-CNN. (a) Grouped five-fold cross-validation accuracies in the representative split (seed = 42). (b) Summary of fusion-model performance across five independent random splits.

(a)Groupedfive-foldcross-validationaccuracy(seed=42)

Fold	LIBSonly	Ramanonly	Fusion
Fold 1	70.00%	83.33%	100.00%
Fold 2	93.33%	86.67%	100.00%
Fold 3	90.00%	76.67%	100.00%
Fold 4	96.67%	93.33%	100.00%
Fold 5	86.67%	90.00%	100.00%
Mean(SD)	**87.33%(10.38%)**	**86.00%(6.41%)**	**100.00%(0.00%)**

(b)Fusion-modelperformanceacrossfiveindependentrandomsplits

Randomseed	CNNfusionCV	CNNfusiontest	SVMfusiontest	RFfusiontest
0	100.00%	100.00%	93.33%	93.33%
1	99.33%	100.00%	93.33%	93.33%
2	100.00%	100.00%	93.33%	90.00%
3	100.00%	96.67%	96.67%	90.00%
42	100.00%	100.00%	96.67%	93.33%
Mean(SD)	**99.87%(0.30%)**	**99.33%(1.49%)**	**94.67%(1.83%)**	**92.00%(1.83%)**

The training curves are provided in Supplementary Fig. S1. Over the last five epochs, the mean training–validation loss difference was ΔL = _0.045 (SD 0.096), and no persistent rise in validation loss was observed. After 5% Gaussian white noise was added to the external- test spectra and the perturbation was repeated 20 times, the fused model still produced an accuracy of 96.00% (SD 2.91%). Together with the repeated-split results, this finding indicates that the single-pulse LIBS strategy did not merely increase input volume; rather, it allowed the model to exploit chemically meaningful spatial heterogeneity that had been averaged out in the shallow-model workflow.

3.4. Deep-feature distribution and model interpretability

The learned feature distributions and Grad-CAM maps are shown in Fig. 3. In the lower-dimensional feature space, samples from different origins exhibited substantially better clustering after dual-stream feature extraction than in the shallow feature representation, indicating that the fused network learned a more compact and better separated latent struc- ture. This result is consistent with the observed performance gain and suggests that the advantage of the dual-stream model comes from its ability to capture nonlinear cross-modal relationships rather than from a simple increase in dimensionality [32, 33].

Table 4: Top five variables highlighted by Grad-CAM in the two branches

(a)LIBSbranch

Wavelength/nm	Assignment	Weight
422.7	Ca I	0.998
589.0	Na I	0.842
279.6	Mg II	0.574
393.3	Ca II	0.536
769.8	K I	0.516

(b)Ramanbranch

Shift/cm__1	Assignment	Weight
1266.1	C–O–C vibration	0.981
1332.8	C–H bending	0.963
1376.5	CH3 bending	0.926
1457.1	CH2 bending	0.856
1117.2	C–C stretching	0.840

Grad-CAM provides a chemically interpretable view of the discriminative process. As summarized in Table 4, the most heavily weighted LIBS variables were concentrated around Ca I (422.7 nm), Na I (589.0 nm), Mg II (279.6 nm), Ca II (393.3 nm), and K I (769.8 nm). On the Raman side, the dominant weights occurred near 1266.1, 1332.8, 1376.5, 1457.1, and

1117.2 cm__1, which are associated with C–O–C, C–H, and C–C vibrations in lignocellulosic structures [30, 31]. Importantly, Grad-CAM does not imply that only these variables mat- ter; rather, it indicates the spectral regions to which the trained network is most sensitive when reaching its decision. Even so, the extracted pattern is chemically coherent: elemen- tal variables dominated a relatively narrow set of discrete LIBS lines, whereas the Raman contribution was distributed across a broader group of structurally informative bands.

3.5. Spectroscopic basis of geographical differences and methodological implications

As summarized in Fig. 4, the geographical discrimination of *Astragali Radix* is not con- trolled by a single elemental line or an isolated Raman band. Instead, it arises from coor- dinated variation at two chemical levels. The LIBS branch records the long-term imprint of mineral supply and root uptake under different soil and ecological conditions, whereas the Raman branch captures structural differences in cellulose-, lignin-, and related cell-wall- associated components. When the highest-contributing variables of the optimal model are considered together, regional origin appears as an integrated elemental–molecular

phenotype rather than as a one-dimensional marker. This point is central to understanding why multi-modal fusion was so effective in the present dataset.

The key variables identified here support that interpretation. The dominant LIBS features— especially Ca, Mg, Na, and K—are chemically plausible indicators of region-dependent min- eral nutrition and rhizosphere conditions. The dominant Raman bands at 1118, 1266, 1336, and 1457 cm__1 correspond to cellulose and lignin frameworks and therefore reflect differences in root structural biochemistry. In practical terms, the two modalities report different con- sequences of the same environmental history: one through elemental accumulation and the other through molecular organization. Their joint use expands the discriminative space be- yond what either elemental or molecular information can provide alone, consistent with the mechanistic framework illustrated in Fig. 4 [16, 17, 18, 19].

This interpretation also clarifies the methodological value of preserving intact-slice het- erogeneity. In conventional workflows, averaging is often necessary to suppress shot-to-shot variation. Here, however, retaining the single-pulse LIBS signals in the deep-learning work- flow allowed local spatial variability to be treated as informative structure rather than as noise. Because the data split was performed at the physical-sample level, the model was not rewarded for memorizing repeated measurements from the same sample across subsets. The performance gain is therefore more reasonably attributed to the effective use of multiscale in- formation: average point-level pairing stabilized the shallow models, whereas the deep model benefited from the richer local variability preserved in the unaveraged LIBS input.

The study nevertheless has clear boundaries. Soil physicochemical properties and absolute concentrations of bioactive compounds were not measured in parallel, so the interpretation of elemental–molecular coordination remains mechanistic rather than directly quantitative. In addition, all samples originated from a single harvest batch for each region, which means that harvest year, storage conditions, and slice thickness still need broader evaluation. Even with these limitations, the repeated sample-level splits showed that the gain from fusion was stable rather than incidental. For intact commercial slices, the present workflow therefore provides evidence that LIBS–Raman fusion can achieve not only high classification accuracy but also a spectroscopically interpretable basis for provenance assignment.

4. Conclusions

This study established a minimally pretreated, in-situ origin-tracing strategy for intact Astragali Radix slices by combining LIBS and Raman spectroscopy. Using 180 samples from six producing regions, we showed that multimodal fusion consistently outperformed single-modality analysis. The fused SVM model achieved a mean external-test accuracy of 94.67% (SD 1.83%), while the fused dual-stream 1D-CNN achieved mean cross-validation and external-test accuracies of 99.87% (SD 0.30%) and 99.33% (SD 1.49%), respectively. The deep model benefited in particular from retaining single-pulse LIBS spectra,

demonstrating that the spatial heterogeneity preserved in intact slices can be converted into useful discriminative information under strict sample-level validation.

Equally important, the classification was chemically interpretable. Grad-CAM linked the fused decisions mainly to Ca-, Mg-, Na-, and K-related LIBS lines and to Raman bands asso- ciated with cellulose and lignin. The geographical signature of Astragali Radix thus appears to be encoded as a coordinated elemental–molecular pattern rather than as a single marker. Taken together, these findings support LIBS–Raman fusion as a high-accuracy and analyt- ically meaningful route for provenance authentication of Astragali Radix slices in practical quality-control settings. Future work should extend the model across harvest years, cultivars, and larger cohorts, and should combine the spectroscopic results with soil properties and quantitative phytochemical data to clarify the mechanistic basis of the most discriminative bands.

Declarations

Supplementary material

Supplementary material associated with this article includes model hyperparameters, the detailed dual-stream network architecture, training curves, and representative external-test confusion matrices.

Funding

This work was supported by the Joint Fund Project of the National Natural Science Foundation of China (U23B2046) and the Science and Technology Program of the Ministry of Public Security of China (2023ZB02).

CRediT authorship contribution statement

Yikang Hou: Conceptualization, Methodology, Investigation, Formal analysis, Software, Visualization, Writing – original draft. **Zhuoxi Li**: Investigation, Data curation, Validation, Writing – review & editing. **Jie Lian**: Conceptualization, Supervision, Funding acquisition, Project administration, Writing – review & editing.

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.

Data availability

Data will be made available on request.

References

1. Z. Chen, L. Liu, C. Gao, et al., Astragali Radix (Huangqi): a promising ed- ible immunomodulatory herbal medicine, J. Ethnopharmacol. 258 (2020) 112895. https://doi.org/10.1016/j.jep.2020.112895.
2. J. Fu, Z. Wang, L. Huang, et al., Review of the botanical characteristics, phytochemistry, and pharmacology of *Astragalus membranaceus* (Huangqi), Phytother. Res. 28 (2014) 1275–1283. https://doi.org/10.1002/ptr.5188.
3. H.-F. Su, S. Shaker, Y. Kuang, et al., Phytochemistry and cardiovascular protec- tive effects of Huang-Qi (Astragali Radix), Med. Res. Rev. 41 (2021) 1999–2038. https://doi.org/10.1002/med.21785.
4. P.Y. Yip, H.S. Kwan, Molecular identification of *Astragalus membranaceus* at the species and locality levels, J. Ethnopharmacol. 106 (2006) 222–229. https://doi.org/10.1016/j.jep.2005.12.033.
5. L. Li, S. Zheng, Q. Yang, et al., Distinguishing *Astragalus mongholicus* and its planting soil samples from different regions by ICP-AES, Molecules 21 (2016) 482. https://doi.org/10.3390/molecules21040482.
6. E.-B. Wang, B.-F. Jin, X. Li, et al., Comparative analysis between aerial parts and roots (Astragali Radix) of *Astragalus membranaceus* by NMR-based metabolomics, Food Agric. Immunol. 28 (2017) 1126–1141. https://doi.org/10.1080/09540105.2017.1332007.
7. P. Dong, L. Wang, D. Qiu, et al., Evaluation of the environmental factors influencing the quality of *Astragalus membranaceus* var. *mongholicus* based on HPLC and the Maxent model, BMC Plant Biol. 24 (2024) 697. https://doi.org/10.1186/s12870-024-05355-3.
8. H. Guo, W. Wang, N. Yang, et al., DNA barcoding provides distinction be- tween Radix Astragali and its adulterants, Sci. China Life Sci. 53 (2010) 992–999. https://doi.org/10.1007/s11427-010-4044-y.
9. L.-W. Qi, Q.-T. Yu, P. Li, et al., Quality evaluation of Radix Astragali through a simultaneous determination of six major active isoflavonoids and four main saponins by high-performance liquid chromatography coupled with diode array and evaporative light scattering detectors, J. Chromatogr. A 1134 (2006) 162–169. https://doi.org/10.1016/j.chroma.2006.08.085.
10. H. Lu, Y. Wang, J. Zhu, et al., Rapid analysis of Radix Astragali using a portable Raman spectrometer with 1064-nm laser excitation and data fusion with PLS-DA, Spectrochim. Acta A Mol. Biomol. Spectrosc. 313 (2024) 124087. https://doi.org/10.1016/j.saa.2024.124087.
11. D.-D. Chen, X.-F. Xie, H. Ao, et al., Raman spectroscopy in quality con- trol of Chinese herbal medicine, J. Chin. Med. Assoc. 80 (2017) 288–296. https://doi.org/10.1016/j.jcma.2016.11.009.
12. Y. Wang, M. Li, T. Feng, et al., Discrimination of Radix Astragali according to geo- graphical regions by data fusion of laser-induced breakdown spectroscopy and infrared spectroscopy combined with random forest, Chin. J. Anal. Chem. 50 (2022) 100057. https://doi.org/10.1016/j.cjac.2022.100057.
13. S. Zhao, W. Song, Z. Hou, Z. Wang, Classification of ginseng according to plant species, geographical origin, and age using laser-induced breakdown spec- troscopy and hyperspectral imaging, J. Anal. At. Spectrom. 36 (2021) 1704–1711. https://doi.org/10.1039/D1JA00136A.
14. F. Zhou, W. Xie, M. Lin, et al., Rapid authentication of geographical origins of Baishao (Radix Paeoniae Alba) slices with laser-induced breakdown spectroscopy based on con- ventional machine

learning and deep learning, Spectrochim. Acta B At. Spectrosc. 212 (2024) 106852. https://doi.org/10.1016/j.sab.2023.106852.

15. E. Borràs, J. Ferré , R. Boqué , et al., Data fusion methodologies for food and beverage authentication and quality assessment – a review, Anal. Chim. Acta 891 (2015) 1–14. https://doi.org/10.1016/j.aca.2015.04.042.
16. C. Eum, D. Jang, J. Kim, et al., Improving the accuracy of spectroscopic identification of geographical origins of agricultural samples through cooperative combination of near- infrared and laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 149 (2018) 281–287. https://doi.org/10.1016/j.sab.2018.09.004.
17. Q. Wang, J. Xiao, Y. Li, et al., Mid-level data fusion of Raman spectroscopy and laser- induced breakdown spectroscopy: improving ores identification accuracy, Anal. Chim. Acta 1240 (2023) 340772. https://doi.org/10.1016/j.aca.2022.340772.
18. L. Ren, F. Wang, Y. Zhang, et al., Rapid identification of Radix Astragali by data fusion of laser-induced breakdown spectroscopy and Raman spectroscopy coupled with deep learning, Talanta 282 (2025) 127016. https://doi.org/10.1016/j.talanta.2024.127016.
19. X. Lin, X. Zhen, J. Lin, et al., Data fusion of laser-induced breakdown spectroscopy (LIBS) and Raman spectroscopy: improving the accuracy of ginseng classification, Anal. Lett. 58 (2025) 1139–1150. https://doi.org/10.1080/00032719.2024.2354904.
20. L.C. Trevizan, D. Santos, R.E. Samad, et al., Evaluation of laser induced breakdown spectroscopy for the determination of macronutrients in plant materials, Spectrochim. Acta B At. Spectrosc. 63 (2008) 1151–1158. https://doi.org/10.1016/j.sab.2008.08.005.
21. L.C. Trevizan, D. Santos, R.E. Samad, et al., Evaluation of laser induced breakdown spectroscopy for the determination of micronutrients in plant materials, Spectrochim. Acta B At. Spectrosc. 64 (2009) 369–377. https://doi.org/10.1016/j.sab.2009.04.003.
22. A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. 36 (1964) 1627–1639. https://doi.org/10.1021/ac60214a047.
23. J. Peng, S. Peng, A. Jiang, et al., Asymmetric least squares for mul- tiple spectra baseline correction, Anal. Chim. Acta 683 (2010) 63–68. https://doi.org/10.1016/j.aca.2010.08.033.
24. Z.-M. Zhang, S. Chen, Y.-Z. Liang, Baseline correction using adaptive it-eratively reweighted penalized least squares, Analyst 135 (2010) 1138–1146.https://doi.org/10.1039/B922045C.
25. C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297. https://doi.org/10.1007/BF00994018.
26. L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32. https://doi.org/10.1023/A:1010933404324.
27. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

28. I. Loshchilov, F. Hutter, Decoupled weight decay regularization, in: 7th Int. Conf. Learn. Represent. (ICLR), New Orleans, LA, USA, 2019. https://doi.org/10.48550/arXiv.1711.05101.
29. R.R. Selvaraju, M. Cogswell, A. Das, et al., Grad-CAM: visual explanations from deep networks via gradient-based localization, in: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, 2017, pp. 618–626. https://doi.org/10.1109/ICCV.2017.74.
30. U.P. Agarwal, J.D. McSweeny, S.A. Ralph, FT-Raman investigation of milled-wood lignins: softwood, hardwood, and chemically modified black spruce lignins, J. Wood Chem. Technol. 31 (2011) 324–344. https://doi.org/10.1080/02773813.2011.562338.
31. U.P. Agarwal, Analysis of cellulose and lignocellulose materials by Raman spectroscopy: a review of the current status, Molecules 24 (2019) 1659. https://doi.org/10.3390/molecules24091659.
32. J.E. Castorena, D.A. Oyen, A.M. Ollila, et al., Deep spectral CNN for laser in- duced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 178 (2021) 106125. https://doi.org/10.1016/j.sab.2021.106125.
33. E. Képeš , J. Vrábel, T. Brázdil, et al., Interpreting convolutional neural network clas- sifiers applied to laser-induced breakdown optical emission spectra, Talanta 266 (2024) 124946. https://doi.org/10.1016/j.talanta.2023.124946.

Figures

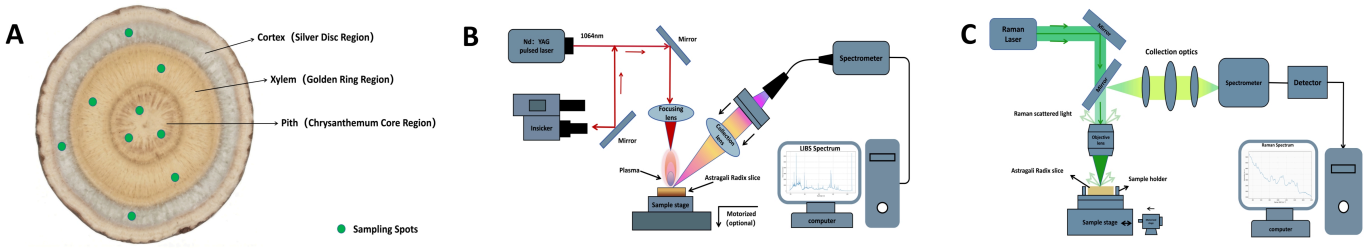

Figure 1

Sample layout and measurement systems. Panel A: Anatomical regions on the Astragali Radix cross section, including the pith, xylem ring, and cortex. Panel B: Schematic diagram of the LIBS

measurement system. Panel C: Schematic diagram of the Raman measurement system.

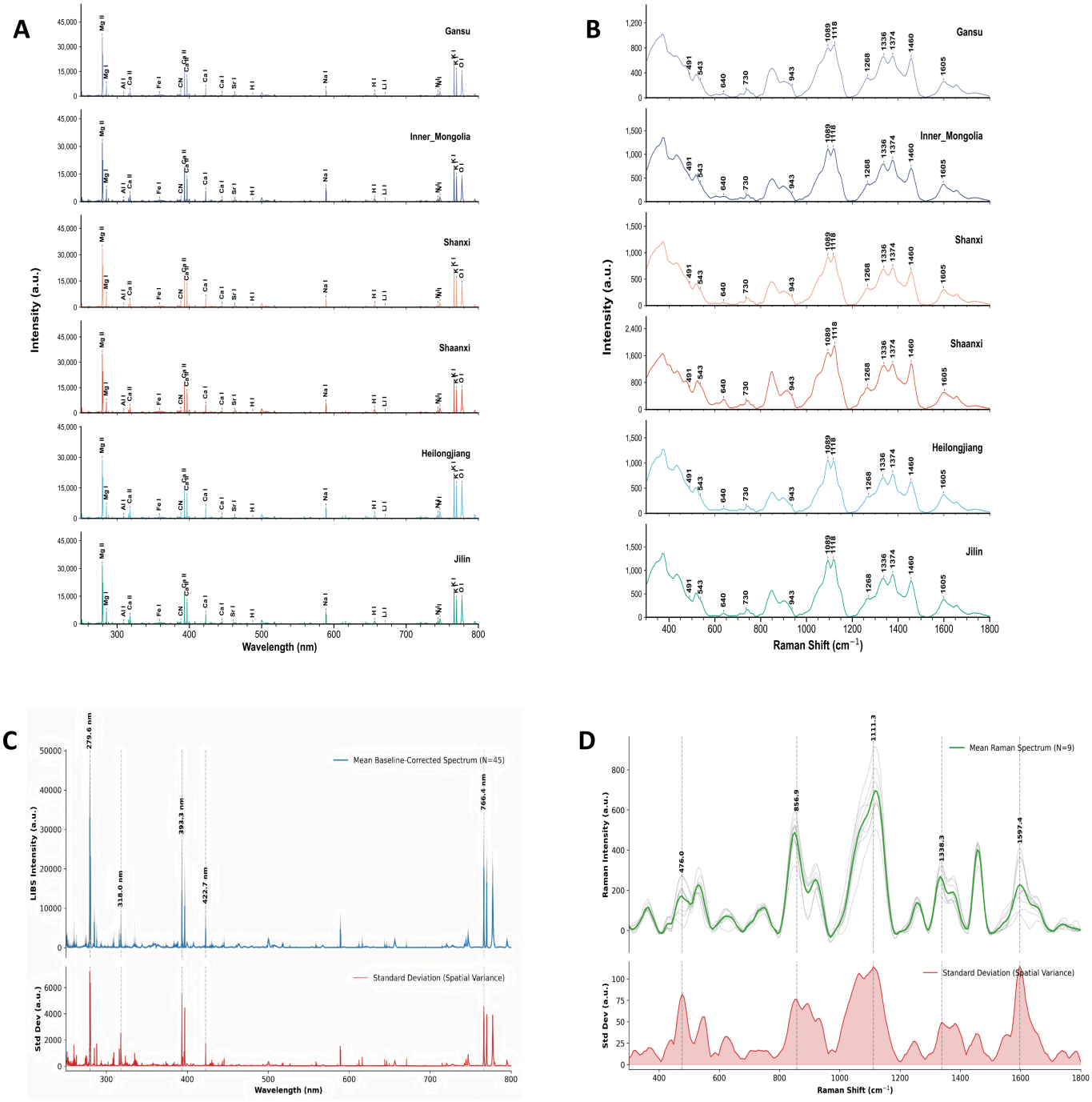

Figure 2

Mean spectra and variance spectra of Astragali Radix from six producing regions.

Panel A: Mean LIBS spectra of samples from the six producing regions. Panel B: Mean Raman spectra of samples from the six producing regions. Panel C: Mean baseline-corrected LIBS spectrum (N = 45) together with the standard deviation representing spatial variance. Panel D: Mean Raman spectrum (N = 9) together with the standard deviation representing spatial variance.

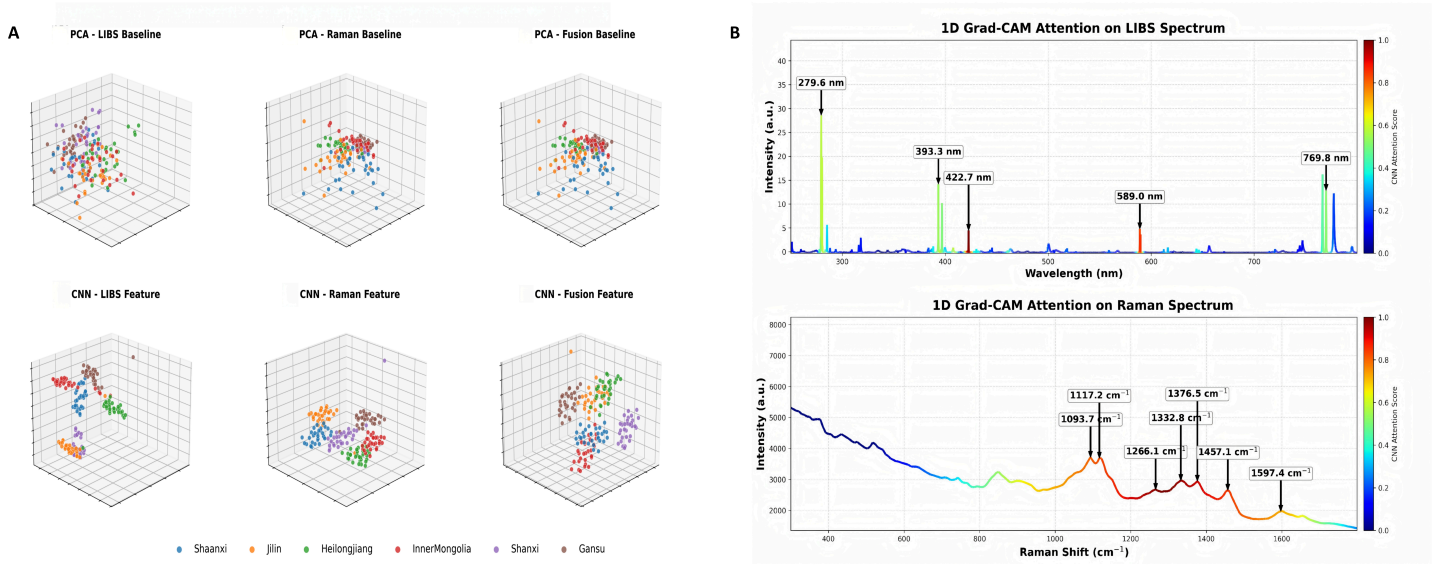

Figure 3

Deep-feature distribution and model interpretability. Panel A: Comparison of baseline PCA feature spaces and learned CNN feature spaces for LIBS, Raman, and fused data. Panel B: One-dimensional Grad-CAM attention maps for the LIBS and Raman branches, highlighting the most discriminative spectral regions used by the model.

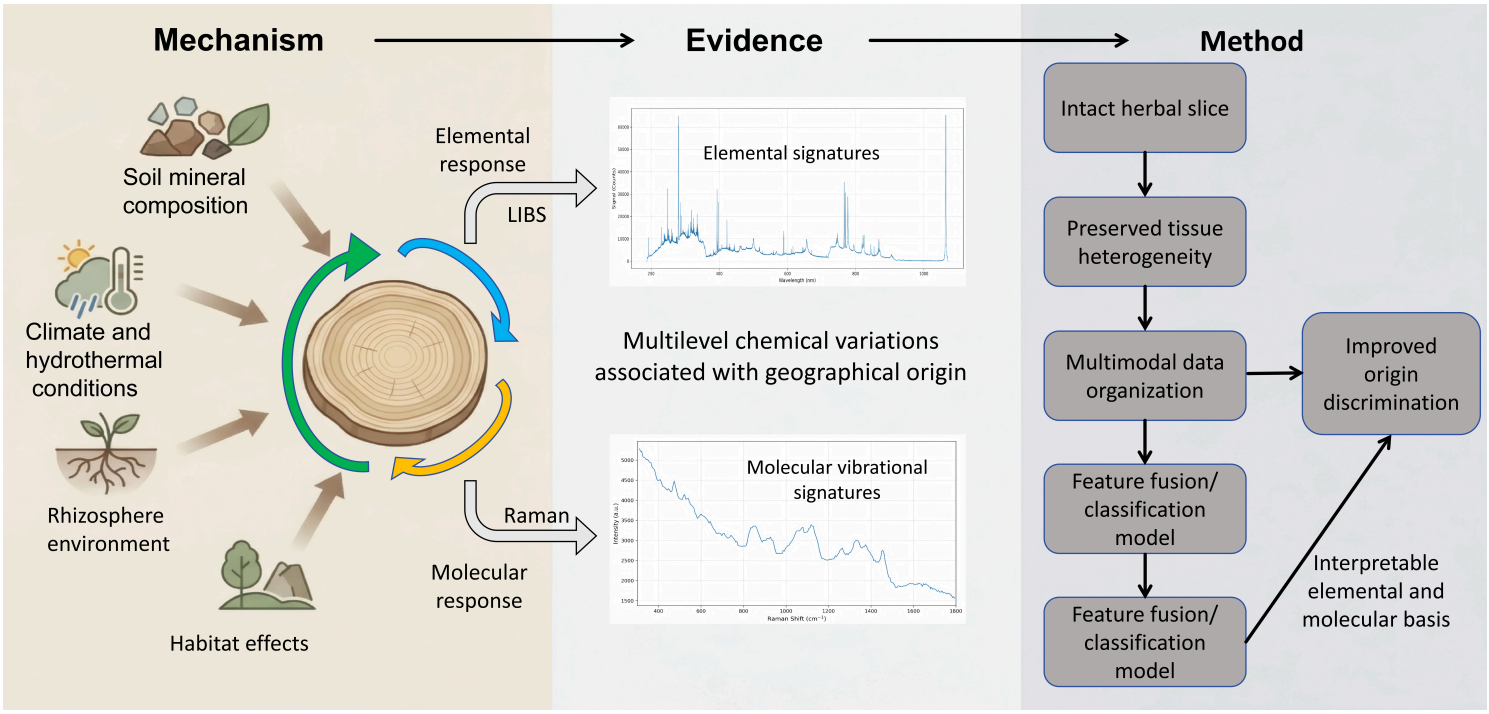

Figure 4

Conceptual summary of the spectroscopic basis of geographical differences and the multimodal analytical workflow. Environmental conditions leave coupled elemental and molecular imprints in the

root; LIBS and Raman record these imprints at different chemical levels; the fused model integrates them for origin classification and interprets the decision through Grad-CAM.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- HouAstragaliRadixOriginTracingSupplementaryMaterial.pdf
- Graphicalabstract.pdf
- FigureS1.jpg
- FigureS2.jpg