Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 25 March 2026
Revised 28 April 2026
Accepted 25 May 2026

Natural Resources for Human Health 2026, Vol. 6 (2s): 569–585

KEYWORDS:
machine learning, chemometrics, herbal medicine, quality control, spectroscopy, deep learning, authentication, adulteration detection, metabolomics

Machine Learning and Chemometrics for Herbal Medicine Quality Control: Analytical Platforms, Performance Benchmarks and Barriers to Regulatory Adoption

Dr. Rajani Sajjan[1], Dr. Jagannath Nalavade[2], Dr. Mukund B. Wagh[3], Dr.Vishnu Suryawanshi[4], Dr. Rajendra G. Pawar[5], Dr. Amar Buchade[6]

[1]Computer Science and Engineering, MIT School of Computing, MIT Art Design and Technology University, Pune 412201, India.
Email: rajani.sajjan@mituniversity.edu.in, Orcid ID: 0000-0002-2679-5206
[2]Computer Science and Engineering, MIT School of Computing, MIT Art Design and Technology University, Pune 412201, India.
Email: jen20074u@gmail.com, Orcid ID: 0000-0002-1645-0095
[3]Artificial Intelligence and Data Science, Maharashtra Institute of Technology, Chhatrapati Sambhajinagar, 431001 India.
Email id: Mukund.wagh.81@gmail.com, Orcid Id:0000-0002-6969-3103
[4]Electrical and Electronics Engineering, MIT School of Computing, MIT Art, Design and Technology University, Pune 412308, India
Email id: vishnusam2007@gmail.com, Orcid Id:0000-0003-0648-2330
[5]Department of Computer Science and Engineering (Artificial Intelligence), Bansilal Ramnath Agarwal Charitable Trust's, Vishwakarma Institute of Technology, Savitribai Phule Pune University, Pune 411037, India,
Email id: rgpawar13@gmail.com, Orcid 0000-0002-8780-3759
[6]Artificial Intelligence and Data Science, BRACT'S Vishwakarma Institute of Technology Pune 411037, India
Email id: amar.buchade@vit.edu, Orcid Id: 0000-0002-6092-6601

ABSTRACT: The global herbal medicine market has expanded rapidly, exceeding an estimated USD 150 billion in valuation, yet quality control remains a critical challenge because of species misidentification, adulteration, and geographical origin fraud. Traditional quality control based on single-marker quantification fails to capture the chemical complexity of herbal matrices. Machine learning and chemometrics have emerged as transformative tools for pattern-based authentication and quality assessment. This review synthesised literature from SciSpace, PubMed, Google Scholar and ArXiv covering 2015–2025, following PRISMA-style selection criteria. Of 1,847 records identified, 450 underwent full-text review and 120 met the inclusion criteria, yielding a final corpus of 150 papers after citation chaining. Papers were classified by analytical platform (vibrational spectroscopy, chromatography–mass spectrometry, hyperspectral imaging, sensor fusion), by chemometric and machine learning method (classical multivariate, supervised learning, deep learning), and by application domain (authentication, adulteration, origin traceability, processing quality, bioactive quantification). Attenuated total reflectance Fourier transform infrared spectroscopy combined with support vector machines achieved 100% classification accuracy for 53 root and rhizome Chinese herbal species. Near-infrared spectroscopy coupled with kernel

extreme learning machines gave 95.6% accuracy in American ginseng adulteration detection. Deep learning architectures, including one-dimensional convolutional neural networks and long short-term memory networks, outperformed classical chemometric methods in spectral feature extraction, with accuracies above 98% reported for hyperspectral imaging-based authentication. Data fusion combining laser-induced breakdown spectroscopy with Raman spectroscopy reached 93.4% accuracy through mid-level fusion, surpassing single-modality approaches. Metabolomics-driven quality marker discovery using backpropagation artificial neural networks yielded correlation coefficients above 0.99 for bioactivity prediction. Machine learning and chemometrics have therefore matured into robust tools that outperform traditional single-marker approaches. Critical gaps nonetheless persist in standardised benchmark datasets, model interpretability for regulatory acceptance, instrument-to-instrument transferability, and translation to portable field devices. Future pathways include explainable artificial intelligence, transfer learning for small-sample scenarios, multi-omics integration, blockchain-enabled supply chain traceability, and edge computing for real-time quality screening.

1. INTRODUCTION

The global herbal medicine market has expanded rapidly over the past decade, driven by consumer preference for natural therapeutics, integration of traditional medicine into mainstream healthcare and growing validation of phytochemical bioactivity. This commercialisation has been accompanied by escalating quality control challenges that threaten both therapeutic efficacy and consumer safety (Jaiswal et al., 2025; Wang et al., 2023; Lam et al., 2023). The chemical composition of botanical materials varies substantially with genetic factors, geographical origin, cultivation practice, harvesting time, post-harvest processing and storage (Liu et al., 2022; Klein-Junior et al., 2021; Kamal, 2022). Species misidentification, whether through morphological similarity or deliberate substitution, is pervasive in global supply chains (Wu and Shaw, 2022; Zhu et al., 2022); adulteration with cheaper botanical material, synthetic compounds or undeclared pharmaceutical ingredients poses serious health risks (Liu et al., 2021); and geographical origin fraud causes economic loss and quality degradation, particularly for herbs carrying protected designation of origin status (Wang and Yu, 2015; Cazelles et al., 2020).

Traditional quality control relies on pharmacopoeial monographs specifying morphological characteristics, microscopic features and quantification of one or two marker compounds by chromatography (Kaur et al., 2026; Singh et al., 2025). This reductionist approach cannot capture the holistic chemical complexity of herbal matrices, in which therapeutic effects often arise from synergistic interactions among multiple constituents (Klein-Junior et al., 2021). Nor can single-marker quantification detect sophisticated adulteration or distinguish closely related species with similar marker profiles (Fernandes et al., 2025; Kamal, 2022).

Machine learning and chemometrics enable pattern-based authentication that leverages the entire chemical fingerprint rather than isolated markers (Abraham and Kellogg, 2021; Wang et al., 2021; Calixto et al., 2021; He and Li, 2024). Supervised algorithms construct classification models that separate authentic samples from adulterants or identify geographical origin with high accuracy (Liu et al., 2025; Zhai et al., 2025; Tang et al., 2022), while deep learning architectures perform automated feature extraction from complex spectral and imaging data (Ji et al., 2024; Bai et al., 2024). Coupling these methods to rapid, non-destructive platforms offers real-time screening in manufacturing and at supply chain checkpoints (Peng et al., 2025; Chen et al., 2023); data fusion across complementary modalities improves discrimination and robustness against instrumental variability (Ding et al., 2023; Ren et al., 2024; Chen et al., 2024); and metabolomics-driven approaches identify quality markers correlating with therapeutic bioactivity (Alum et al., 2025; Yang et al., 2021).

This review synthesises a decade of research (2015–2025) with three objectives: to categorise the analytical platforms, chemometric methods and machine learning algorithms employed; to evaluate performance benchmarks across application domains; and to identify research gaps and pathways towards regulatory acceptance and field deployment. Section 2 sets out the quality control problem and its regulatory context, Section 3 the review methodology, Sections 4 and 5 the analytical

platforms and computational methods, Section 6 the applications, Section 7 a comparative analysis, and Sections 8 to 10 the challenges, future directions and conclusions.

2. HERBAL QUALITY CONTROL: CHALLENGES AND REGULATORY CONTEXT

Unlike synthetic pharmaceuticals with defined molecular structures, herbal medicines are complex mixtures of hundreds to thousands of constituents whose composition varies with biological and environmental factors (Wang et al., 2023; Klein-Junior et al., 2021). Regulatory frameworks differ substantially across jurisdictions. The European Medicines Agency requires demonstration of traditional or well-established medicinal use with standards defined in European Pharmacopoeia monographs; the United States Food and Drug Administration regulates most herbal products as dietary supplements with less stringent pre-market requirements; and the China National Medical Products Administration enforces comprehensive standards through the Chinese Pharmacopoeia, specifying marker compounds, chromatographic fingerprints and biological assays (Durazzo et al., 2022; Lindenmaier et al., 2025).

Five scientific challenges recur. Species authentication is complicated by morphological similarity among related taxa, particularly in processed forms where diagnostic features are destroyed (Liu et al., 2022; Kamal, 2022; Wu and Shaw, 2022; Zhu et al., 2022). Adulteration ranges from substitution with cheaper species to addition of synthetic compounds or heavy metals (Liu et al., 2021). Geographical origin influences phytochemical composition and therefore requires provenance verification (Wang and Yu, 2015; Cazelles et al., 2020). Processing by steaming, stir-frying or fermentation alters chemical profiles and bioactivity, so quality control must distinguish processing states (Ji et al., 2024). Bioactive quantification must account for synergy and identify markers correlating with efficacy rather than arbitrary constituents (Yang et al., 2021; Fernandes et al., 2025). Pattern-based approaches capturing holistic fingerprints address these challenges more effectively than morphological examination or single-marker quantification (Wang et al., 2021; He and Li, 2024; Klein-Junior et al., 2021).

3. REVIEW METHODOLOGY

Searches were conducted across SciSpace, PubMed, Google Scholar and ArXiv for January 2015 to December 2025, combining controlled vocabulary and free-text terms across three concept domains: herbal medicine terms; analytical platform terms including spectroscopy, chromatography, mass spectrometry and hyperspectral imaging; and computational method terms including machine learning, chemometrics, deep learning and pattern recognition. Studies were included if they addressed quality control, authentication or standardisation of herbal medicines, employed chemometric or machine learning methods on instrumental data, reported quantitative performance metrics, and were published in peer-reviewed journals or preprint repositories in English. Studies addressing pharmacological efficacy without a quality control application, employing only morphological methods, or lacking quantitative evaluation were excluded, as were non-herbal natural products unless the methodology was directly transferable.

Database searches yielded 1,847 records; 1,203 remained after duplicate removal, 450 underwent full-text review and 120 met the inclusion criteria, with 30 further papers identified through citation chaining, giving a final corpus of 150 (Figure 1). Papers were classified along three dimensions — analytical platform, computational method and application domain (Figure 2). Extraction covered species, sample size and origin, acquisition parameters, preprocessing, algorithm, validation strategy and performance metrics, and was performed by two independent reviewers with discrepancies resolved by consensus. Methodological quality was assessed using criteria adapted from PRISMA and STARD covering sample size adequacy, representativeness of sampling, appropriateness of validation, transparency of preprocessing and completeness of performance reporting.

Natural Resources for Human Health
eISSN: 2583-1194

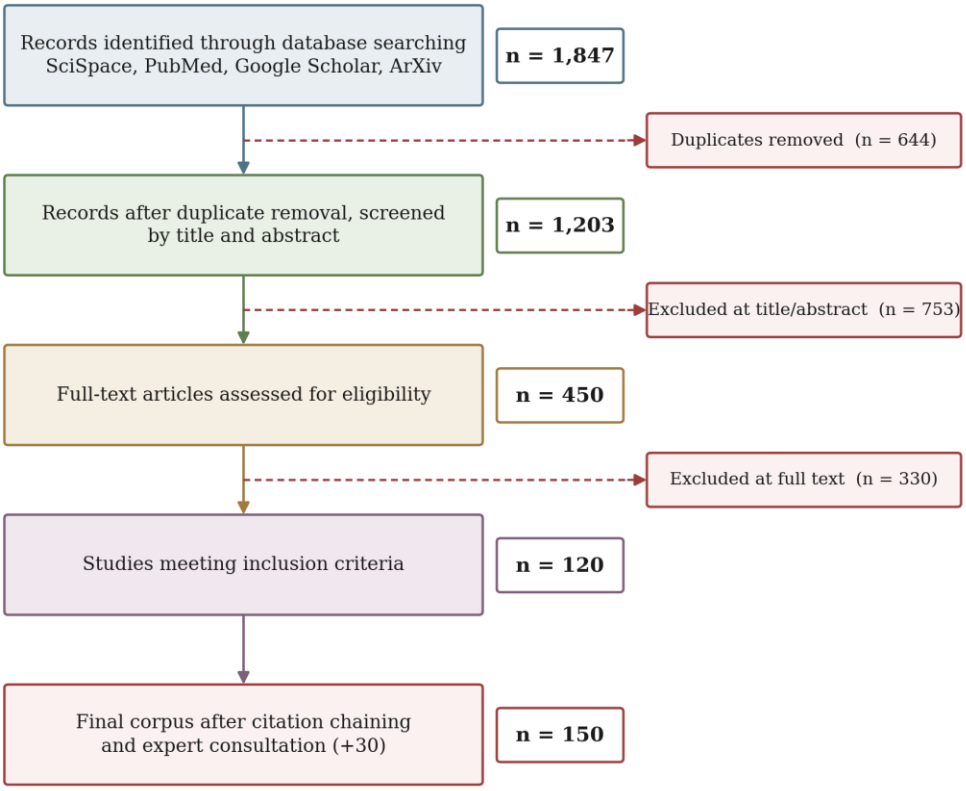

Figure 1. Flow of study identification, screening, eligibility assessment and inclusion, following PRISMA-style criteria. Counts correspond to the search described in Section 3.

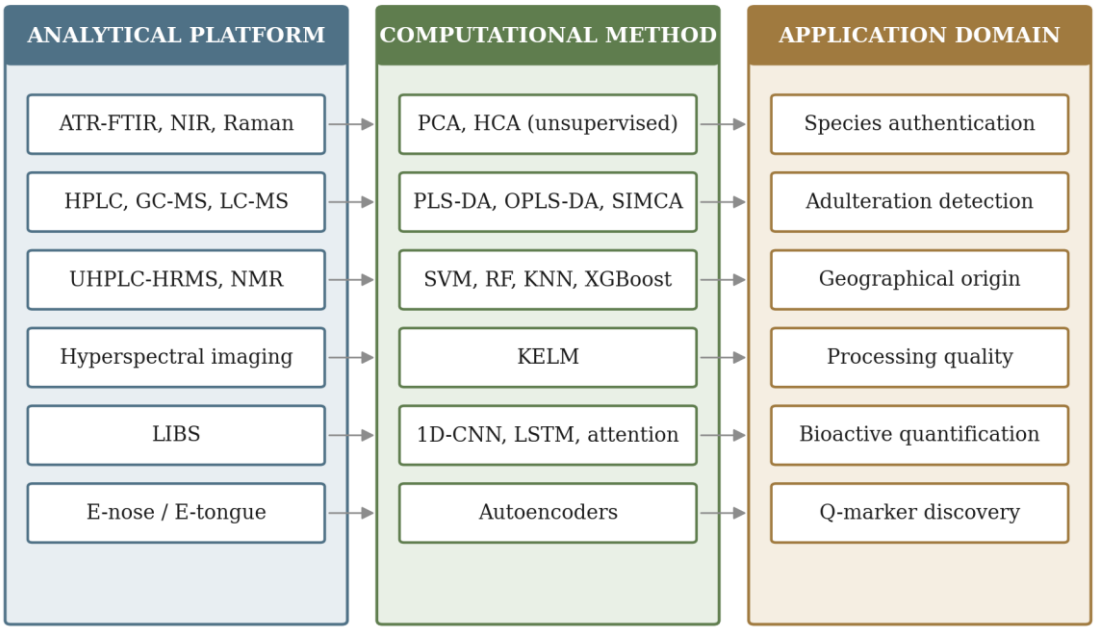

Figure 2. Three-axis framework used to classify the reviewed literature. Each study is positioned by one selection from each axis, which allows platform, method and application to be compared independently.

4. ANALYTICAL PLATFORMS AND DATA ACQUISITION

The platforms encountered differ in the chemical information captured, sample preparation burden and suitability for field deployment; their characteristics are compared in Table 1.

Table 1. Analytical platforms applied to herbal quality control, with their principal characteristics.

Platform	Information captured	Principal strengths	Principal limitations
ATR-FTIR	Mid-infrared molecular vibrations; functional groups	Minimal preparation; rich chemical detail; high throughput	Surface-sensitive; water interference
NIR	Overtone and combination bands	Intact samples; handheld instruments; field-deployable	Broad overlapping bands; needs chemometric calibration
Raman	Inelastic scattering; symmetric and skeletal modes	Water-insensitive; complementary to infrared	Fluorescence interference; weak signal
HPLC / GC-MS / LC-MS	Separated constituent profiles; molecular masses	High specificity; targeted quantification; metabolomics	Destructive; slow; solvent-intensive
NMR	Quantitative metabolite profile	No separation required; inherently quantitative	Low sensitivity; high instrument cost
Hyperspectral imaging	Spatial plus spectral data cube	Detects localised adulteration and heterogeneity	Large data volume; specialised preprocessing
LIBS	Atomic emission; elemental composition	Rapid multi-element analysis; origin markers	Micro-destructive; matrix effects
E-nose / E-tongue	Volatile and electrochemical response	Low cost; rapid; complements spectroscopy	Sensor drift; limited chemical specificity

4.1 Vibrational Spectroscopy

Vibrational methods probe molecular vibrations to generate spectral fingerprints, offering rapid non-destructive analysis with minimal preparation (Chen et al., 2023). ATR-FTIR measures mid-infrared absorption (4000–400 cm^{-1}) and, combined with optimised SVM, achieved 100% classification accuracy for 53 root and rhizome Chinese herbal species with substantially shorter identification time than PLS-DA (Liu et al., 2025). It has also detected metanil yellow adulteration in turmeric at 1% and authenticated *Dendrobium* species using OPLS-DA (Noviana et al., 2022; Zhao et al., 2025; Hao et al., 2021).

NIR spectroscopy exploits overtone and combination bands in the 780–2500 nm region and analyses intact samples without grinding, with handheld instruments enabling field deployment. Random forest with SMOTETomek resampling achieved area under the curve values of 0.997 in cross-validation and 0.945 on test data for Cuscutae Semen authentication (Peng et al., 2025); a kernel extreme learning machine gave 95.6% accuracy for American ginseng adulteration (Zhai et al., 2025); and Bayesian-optimised LSTM networks achieved 100% classification accuracy with R^2 above 0.99 for adulteration quantification in Angelicae Sinensis Radix (Bai et al., 2024). Raman spectroscopy measures inelastic scattering, is insensitive to water interference, and has enabled ratio quantification of essential oil mixtures (Tang et al., 2025), detection of adulterated Suichang native honey (Hu et al., 2022) and authentication of Radix Astragali when fused with LIBS (Ren et al., 2024). All three platforms generate high-dimensional data requiring preprocessing by scatter correction, standard normal variate transformation, smoothing and derivative spectroscopy (Peng et al., 2025; Chen et al., 2023).

4.2 Chromatographic and Mass Spectrometric Methods

Chromatographic separation with spectroscopic or mass spectrometric detection provides detailed profiling for both targeted quantification and untargeted metabolomics (Alum et al., 2025; Riswanto et al., 2022). HPLC-UV fingerprints coupled with PLS-DA achieved R^2X of 0.965, R^2Y of 0.958 and Q^2(cum) of 0.93 for *Curcuma xanthorrhiza* adulteration detection (Kusumadewi

et al., 2022). Headspace GC-MS with a flash GC electronic nose and XGBoost achieved 86.2% accuracy for *Atractylodes lancea* variety, origin and production mode (Gan et al., 2024). UPLC-LTQ-Orbitrap metabolomics with a backpropagation neural network achieved R above 0.9983 for quality marker discovery (Yang et al., 2021). ^{1}H-NMR provides quantitative profiling without separation and, with OPLS-DA, achieved R^2X of 0.977, R^2Y of 0.962 and Q^2 of 0.928 for black pepper origin discrimination while also detecting artemisinin adulteration in *Artemisia afra* capsules (Riswanto et al., 2022; Rebiai et al., 2022).

4.3 Imaging, Elemental and Multi-Modal Platforms

Hyperspectral imaging acquires hundreds of contiguous bands per pixel, enabling simultaneous assessment of composition and spatial distribution and thereby detecting localised adulteration, foreign matter and heterogeneity that bulk analysis would miss (Filho et al., 2023; Kabir et al., 2022b). Deep learning with HSI achieved 97.6% accuracy for *Fritillaria thunbergii* variety discrimination (Kabir et al., 2022b), while PLS-DA gave above 95% accuracy for medicinal plant species identification and SVM with feature selection reached 100% validation accuracy for *Astragalus membranaceus* seed authentication (Filho et al., 2023).

LIBS ablates sample material with a focused laser pulse and reads the plasma emission spectrum, providing rapid multi-element analysis suited to heavy metal detection and origin discrimination from mineral profiles (Kabir et al., 2022a; Zhou et al., 2024b). Fusion strategies combine platforms at low, mid or high level (Ding et al., 2023; Ren et al., 2024; Chen et al., 2024). Mid-level fusion of LIBS and Raman achieved 93.4% accuracy for Radix Astragali, outperforming low-level fusion at 88.3% and high-level fusion at 90.1% (Ren et al., 2024), and deep learning fusion of LIBS with visible–near-infrared spectroscopy reached 98.75% for Polygonati Rhizoma adulteration (Chen et al., 2024). Multi-spectra fusion with Bayesian-optimised learning outperformed single-modality approaches for *Panax notoginseng* (Guan et al., 2024). Electronic tongue and nose sensors add electrochemical and olfactory information, explaining 100% of data variation in one quality control study (Taha and Abu-Khalaf, 2020) and distinguishing decoction taste profiles with 88.2% accuracy (Liang and Liu, 2025).

5. CHEMOMETRIC AND MACHINE LEARNING METHODS

Table 2. Computational methods applied in herbal quality control, with representative reported performance.

Family	Method	Characteristic	Representative reported performance
Classical chemometrics	PCA	Unsupervised variance projection	100% of data variation explained, e-tongue (Taha and Abu-Khalaf, 2020)
Classical chemometrics	PLS-DA	Supervised latent variables	R^2X 0.965, R^2Y 0.958, Q^2 0.93, Curcuma (Kusumadewi et al., 2022)
Classical chemometrics	OPLS-DA	Separates orthogonal variation	R^2X 0.977, R^2Y 0.962, Q^2 0.928, black pepper (Riswanto et al., 2022)
Classical chemometrics	HCA	Similarity-based dendrograms	Reliable quality control, e-tongue (Taha and Abu-Khalaf, 2020)
Classical chemometrics	SIMCA	One-class distance modelling	Four Actaea species at 95% confidence (Durazzo et al., 2022)
Supervised ML	SVM	Margin-maximising hyperplanes	100% accuracy, 53 species, ATR-FTIR (Liu et al., 2025)
Supervised ML	Random forest	Bagged decision-tree ensemble	AUC 0.997 CV / 0.945 test, Cuscutae Semen (Peng et al., 2025)
Supervised ML	KNN	Distance-based voting	Competitive across authentication studies (Liu et al., 2025; Riswanto et al., 2022)
Supervised ML	XGBoost	Gradient boosting ensemble	86.2% accuracy, Atractylodes lancea (Gan et al., 2024)

Family	Method	Characteristic	Representative reported performance
Supervised ML	KELM	Analytic single-layer solution	95.6% accuracy, Kappa 0.955, ginseng (Zhai et al., 2025)
Deep learning	1D-CNN	Learned local spectral features	98.75% accuracy, LIBS–VNIR fusion (Chen et al., 2024)
Deep learning	Attention 1D-CNN	Weighted informative regions	98.97% accuracy, Fritillaria HSI (Kabir et al., 2022b)
Deep learning	LSTM	Long-range sequential dependencies	100% classification, $R^2 > 0.99$, NIR (Bai et al., 2024)
Deep learning	Autoencoder	One-class anomaly detection	Raman pharmaceutical screening (Somov, 2025)

5.1 Classical Multivariate Methods

Classical chemometric methods provide interpretable models with transparent decision boundaries, assuming linear or low-order relationships between features and quality attributes (He and Li, 2024). PCA projects data onto orthogonal components capturing maximum variance and explained 100% of variation in electronic tongue-based quality control (Taha and Abu-Khalaf, 2020); combined with OPLS-DA it differentiated *Curcuma* species with R^2 above 0.8 and Q^2 above 0.45 (Noviana et al., 2022). PLS-DA constructs latent variables maximising covariance between spectral features and class labels, achieving above 95% accuracy for HSI-based species identification (Filho et al., 2023) and R^2X of 0.965 with Q^2(cum) of 0.93 for *Curcuma xanthorrhiza* (Kusumadewi et al., 2022). OPLS-DA separates systematic from orthogonal variation, achieving Q^2 of 0.928 for black pepper origin (Riswanto et al., 2022) and enabling rapid *Dendrobium* authentication with ATR-FTIR and FT-NIR (Zhao et al., 2025). HCA constructs dendrograms visualising hierarchical relationships (Taha and Abu-Khalaf, 2020), while SIMCA models each class separately and discriminated four *Actaea* species at the 95% confidence limit (Durazzo et al., 2022). These methods are computationally efficient and robust with small samples but may underperform on highly nonlinear patterns (Wang et al., 2021; Joshi, 2023).

5.2 Supervised Machine Learning

Supervised algorithms learn classification functions from labelled data and often outperform classical chemometrics on complex tasks (Liu et al., 2025; Wang et al., 2021; Tang et al., 2022). SVM constructs margin-maximising hyperplanes and achieved 100% accuracy for ATR-FTIR authentication of 53 species (Liu et al., 2025), 96.2% recognition accuracy for infrared-based identification (Tang et al., 2022) and 100% validation accuracy for HSI-based Astragalus seed authentication (Filho et al., 2023). Random forests aggregate bootstrap-trained decision trees, achieving area under the curve values of 0.997 and 0.945 for handheld NIR-based Cuscutae Semen authentication (Peng et al., 2025) and competitive performance for GC-MS discrimination (Gan et al., 2024). KNN achieves competitive results when paired with appropriate distance metrics (Riswanto et al., 2022; Zhai et al., 2025). XGBoost reached 86.2% for *Atractylodes lancea* (Gan et al., 2024), and KELM achieved 95.6% accuracy with a Kappa coefficient of 0.955 for NIR-based ginseng adulteration (Zhai et al., 2025). These algorithms nonetheless function largely as black boxes, which may hinder regulatory acceptance (Vellido, 2020).

5.3 Deep Learning Architectures

Deep learning learns hierarchical representations automatically, removing the need for manual feature engineering (Ji et al., 2024). One-dimensional CNNs extract local spectral patterns and achieved 98.75% accuracy for LIBS–VNIR fusion (Chen et al., 2024), while attention-enhanced 1D-CNN reached 98.97% for HSI-based *Fritillaria* authentication (Kabir et al., 2022b). LSTM networks capture long-range dependencies and, under Bayesian optimisation, achieved 100% classification accuracy with R^2 above 0.99 for NIR-based quality evaluation (Bai et al., 2024). CNNs applied to hyperspectral images learn spatial and spectral features simultaneously, enabling non-destructive detection of red pepper powder adulteration (Choi et al., 2025). One-class autoencoders permit anomaly detection without labelled adulterant samples (Somov, 2025), and few-shot class incremental

learning enabled chrysanthemum identification from only 30 samples per new class (Filho et al., 2023). These gains come at the cost of large training sets, substantial computation and careful hyperparameter tuning (Ji et al., 2024; Bai et al., 2024).

5.4 Data Fusion Strategies

Low-level fusion concatenates raw spectra before feature extraction and is computationally simple but suffers from high dimensionality and unequal modality weighting; mid-level fusion concatenates separately extracted features and balances integration against dimensionality; and high-level fusion combines per-modality classifier outputs, providing robustness against modality-specific failure while potentially discarding complementary information (Ding et al., 2023; Ren et al., 2024). For Radix Astragali these gave 88.3%, 93.4% and 90.1% respectively (Figure 3b). Deep fusion networks learn the fusion strategy end to end: LVDLNet, combining LIBS and VNIR spectroscopy, achieved 98.75% accuracy with 98.5% macro-F1 (Chen et al., 2024).

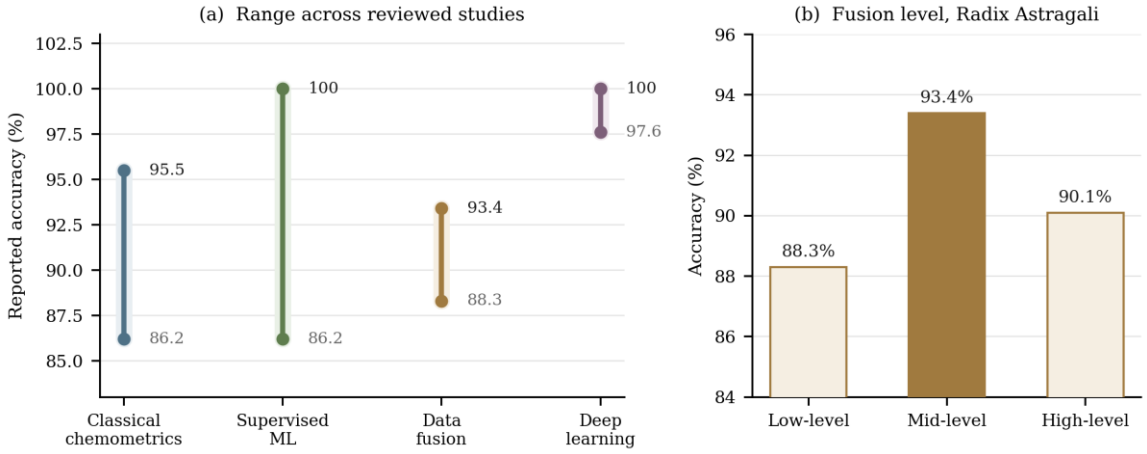

Figure 3. (a) Range of reported accuracy across the reviewed studies for each family of computational method. (b) Effect of data fusion level on identification accuracy for Radix Astragali, after Ren et al. (2024).

6. APPLICATIONS IN HERBAL QUALITY CONTROL

6.1 Species Authentication

ATR-FTIR with optimised SVM achieved 100% accuracy for 53 root and rhizome species, outperforming PLS-DA, decision trees, naive Bayes, KNN and ensemble learning in both accuracy and speed (Liu et al., 2025). Infrared spectroscopy with improved linear discriminant analysis achieved 96.2% recognition accuracy (Tang et al., 2022), hyperspectral imaging with PLS-DA exceeded 95% for cerrado medicinal plants (Filho et al., 2023), and deep learning image recognition reached 99.0% accuracy with an area under the curve of 0.9942 across 163 traditional Chinese medicine types (Liang and Liu, 2025). Notably, cluster analysis differentiated four *Bupleurum* species with 88.6% accuracy against 67.7% for a CNN, showing that classical approaches remain competitive for some tasks (Bai and Zhang, 2024). Across these studies, pattern-based approaches consistently outperformed single-marker quantification for closely related species (Liu et al., 2022; Kamal, 2022).

6.2 Adulteration Detection and Quantification

NIR with KELM achieved 95.6% accuracy and a Kappa coefficient of 0.955 for American ginseng powder, outperforming KNN, decision trees and random forests (Zhai et al., 2025). A handheld NIR spectrometer with random forest and SMOTETomek addressed imbalanced categories, reaching area under the curve values of 0.995 and 1.000 for content authenticity with f-score and recall of 1.000 (Peng et al., 2025). ATR-FTIR detected metanil yellow in turmeric at 1%, and PLSR models for *Fritillaria unibracteata* adulteration achieved R^2 of 0.9072 with RMSEP of 0.1004 (Noviana et al., 2022). Deep learning fusion reached 98.75% for Polygonati Rhizoma (Chen et al., 2024), while Raman with mid-infrared spectroscopy detected adulterants in ground turmeric (Teklemariam, 2024) and multi-spectra fusion performed strongly for *Panax notoginseng* (Guan et al., 2024). Fusion and deep learning therefore add most value for subtle adulteration, while classical methods remain adequate for gross substitution (Liu et al., 2021).

6.3 Geographical Origin and Traceability

^{1}H-NMR with OPLS-DA achieved Q^2 of 0.928 for black pepper origin, capturing region-specific signatures reflecting soil, climate and cultivation (Riswanto et al., 2022). LIBS elemental profiles reflect soil geochemistry and enabled rapid authentication of Baishao origin, with deep learning outperforming conventional machine learning (Zhou et al., 2024b). NIR with artificial intelligence algorithms achieved 96.4% accuracy for origin identification of authentic medicinal materials (Liang and Liu, 2025) and has been applied to *Gastrodia elata* and related species (Yang et al., 2023). Hyperspectral imaging added within-sample heterogeneity detection (Filho et al., 2023), and GC-MS with electronic nose and XGBoost achieved 86.2% for *Atractylodes lancea*, showing that volatile profiles complement non-volatile metabolites (Gan et al., 2024). Platform choice should therefore match the chemical signature that distinguishes the origins of interest (Wang and Yu, 2015; Cazelles et al., 2020).

6.4 Processing and Storage Quality

NIR with deep learning discriminated processed traditional Chinese medicine products, capturing processing-induced chemical transformation (Ji et al., 2024). Electronic tongue sensors with PCA and HCA identified ingredient changes and monitored shelf life, with standard deviations of 5.6–10.4 mV and coefficients of variation of 0.31–0.61% (Taha and Abu-Khalaf, 2020). HPLC fingerprints distinguished raw from processed Cuscutae Semen, Mume Fructus and *Panax notoginseng*, and metabolomics identified compounds that increase or decrease during processing (Riswanto et al., 2022). FT-NIR with multiple intelligent algorithms enabled rapid evaluation of *Smilax glabra* Roxb. quality (Zhan et al., 2024).

6.5 Bioactive Quantification and Quality Marker Discovery

UPLC-LTQ-Orbitrap metabolomics with a backpropagation neural network achieved R above 0.9983 and mean squared error below 0.0014 for quality marker discovery in the Jinqi Jiangtang preparation against type 2 diabetes, with ReliefF feature selection identifying markers associated with antidiabetic bioactivity and thereby grounding standardisation in therapeutic relevance rather than arbitrary constituents (Yang et al., 2021; Fernandes et al., 2025). Artificial intelligence-based standard curve prediction achieved 95.67% accuracy against 90.67% for traditional methods (Mall et al., 2025). FT-NIR assisted learning determined acteoside, aucubin and catalpol contents in *Plantago lanceolata* (Camps, 2020a; Camps et al., 2020b), and NIR with deep learning quantified active ingredients in Radix Gentianae Macrophyllae extracts (Peng et al., 2024). NIR models for total polyphenols in tea achieved R^2 of 0.994 with RMSEP of 0.595 (Arifah et al., 2022), and PLSR models for curcuminoids exceeded R^2 of 0.99 (Kusumadewi et al., 2022), matching or exceeding chromatographic quantification while being faster and non-destructive.

6.6 Metabolomics-Driven Standardisation

Comprehensive profiling captures holistic composition and supports quality control accounting for multi-component synergy, using NMR, LC-MS, GC-MS, DESI-MS and DART-MS with PCA, PLS-DA and machine learning (Alum et al., 2025). NMR metabolomics detected falsely present artemisinin in *Artemisia afra* capsules, distinguished *A. afra* from *A. annua*, and differentiated cultivation sources of *Panax ginseng* (Rebiai et al., 2022). LC-MS metabolomics profiled species- and variety-specific variation in *Ocimum*, while integration of untargeted with pseudotargeted metabolomics revealed authentication markers for *Fritillaria*e Bulbus and enabled quality assessment of *Aconitum heterophyllum* (Abraham and Kellogg, 2021).

7. COMPARATIVE ANALYSIS AND PERFORMANCE BENCHMARKS

Representative studies with their reported metrics are summarised in Table 3, with accuracies grouped by method family in Figure 4. Vibrational spectroscopy with machine learning consistently exceeded 95% for authentication and adulteration detection (Liu et al., 2025; Peng et al., 2025; Zhai et al., 2025; Tang et al., 2022). Chromatographic and mass spectrometric methods offered superior specificity for marker identification, with OPLS-DA models on NMR and LC-MS data exceeding R^2 and Q^2 of 0.9 (Riswanto et al., 2022; Kusumadewi et al., 2022; Yang et al., 2021). Hyperspectral imaging with deep learning approached 98–99% while adding spatial resolution (Filho et al., 2023; Kabir et al., 2022b).

Among computational methods, classical chemometrics performed robustly with small samples and transparent models (Bai and Zhang, 2024; Kusumadewi et al., 2022); supervised learning generally outperformed it, with SVM reaching 100% for large-scale authentication (Liu et al., 2025) and ensembles providing built-in feature importance (Peng et al., 2025; Gan et al., 2024); and deep learning achieved the highest reported accuracies, with Bayesian-optimised LSTM reaching 100% classification and R^2 above 0.99 (Bai et al., 2024; Chen et al., 2024; Kabir et al., 2022b). Figure 3a shows the range each family spans. By domain,

authentication consistently exceeded 95% given adequate sampling (Liu et al., 2025; Liang and Liu, 2025; Tang et al., 2022); adulteration detection varied with adulterant type, gross substitution exceeding 95% while subtle cases required fusion or deep learning (Peng et al., 2025; Zhai et al., 2025; Chen et al., 2024); origin discrimination exceeded R^2 and Q^2 of 0.9 for chemically distinct origins but degraded for closely related ones (Riswanto et al., 2022; Gan et al., 2024; Zhou et al., 2024b); and bioactive quantification exceeded R^2 of 0.99 for well-calibrated models (Kusumadewi et al., 2022; Yang et al., 2021; Arifah et al., 2022).

The trade-offs are systematic. Classical chemometrics offers efficiency, transparency and small-sample robustness but underperforms on nonlinear patterns (Wang et al., 2021; He and Li, 2024; Joshi, 2023). Supervised learning improves accuracy but needs larger training sets and careful validation (Liu et al., 2025; Zhai et al., 2025). Deep learning performs best on high-dimensional data but demands computation and tuning (Bai et al., 2024; Chen et al., 2024; Kabir et al., 2022b). Single-modality approaches are simpler and cheaper but less robust to instrumental variability, whereas fusion improves discrimination at the cost of complexity (Ding et al., 2023; Ren et al., 2024).

Table 3. Performance benchmarks of representative studies. All values are as reported by the original authors.

Study	Platform	Method	Application	Reported performance
Liu et al. (2025)	ATR-FTIR	SVM	53 species authentication	Accuracy 100%
Peng et al. (2025)	Handheld NIR	RF + SMOTETomek	Cuscutae Semen authentication	AUC 0.997 CV, 0.945 test; F1 0.954
Zhai et al. (2025)	NIR	KELM	Ginseng adulteration	Accuracy 95.6%; Kappa 0.955
Bai and Zhang (2024)	ATR-FTIR	Cluster analysis	Bupleurum differentiation	Accuracy 88.6%; sensitivity 86.5%
Riswanto et al. (2022)	^{1}H-NMR	OPLS-DA	Black pepper origin	R^2X 0.977; R^2Y 0.962; Q^2 0.928
Kusumadewi et al. (2022)	FTIR, HPLC	PLS-DA, PLSR	Curcuma authentication	Accuracy 95.5%; $R^2 > 0.99$
Yang et al. (2021)	UPLC-Orbitrap	BP-ANN	Q-marker discovery	R 0.9983; MSE < 0.0014
Bai et al. (2024)	NIR	Bayesian-optimised LSTM	Angelicae Sinensis quality	Classification 100%; $R^2 > 0.99$
Chen et al. (2024)	LIBS + VNIR	LVDLNet deep fusion	Polygonati Rhizoma adulteration	Accuracy 98.75%; macro-F1 98.5%
Ren et al. (2024)	LIBS + Raman	Mid-level fusion + CNN	Radix Astragali identification	Accuracy 93.4%
Kabir et al. (2022b)	HSI	Attention 1D-CNN	Fritillaria discrimination	Binary 98.97%; 8-class 97.64%
Liang and Liu (2025)	Digital imaging	CNN	163 TCM types	Accuracy 99.0%; AUC 0.9942
Gan et al. (2024)	GC-MS + e-nose	XGBoost	Atractylodes variety and origin	Accuracy 86.2 ± 7.5%
Noviana et al. (2022)	ATR-FTIR	Multivariate	Turmeric adulteration	Detection limit 1% metanil yellow
Arifah et al. (2022)	UV-Vis, NIR	ANN, PLSR	Tea authentication	Classification 100%; R^2 0.994

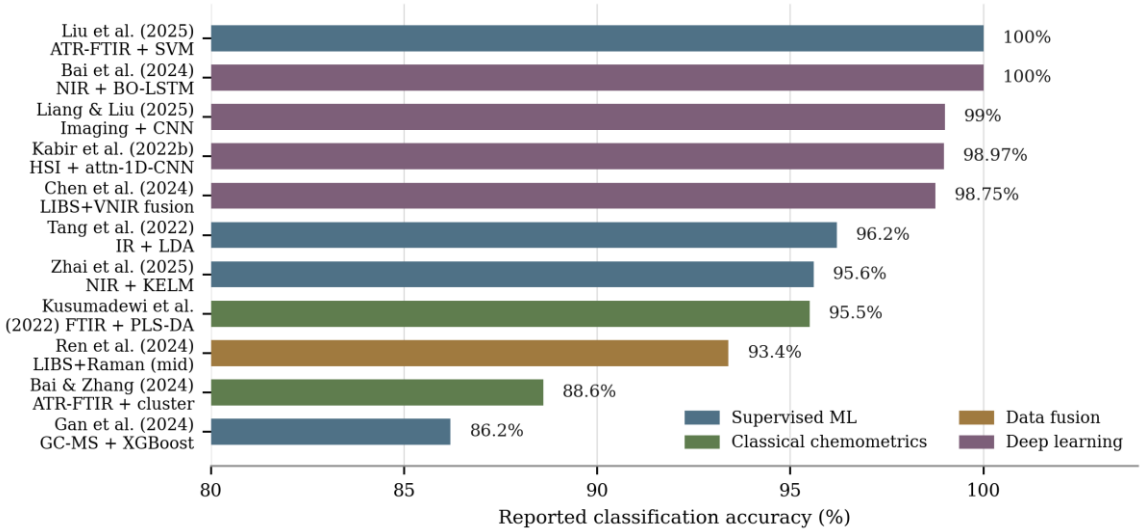

Figure 4. Classification accuracy reported by representative studies, grouped by the family of computational method employed. All values are those stated by the original authors and are not recomputed here.

8. CHALLENGES AND RESEARCH GAPS

Table 4 summarises the principal challenges and recommended responses, mapped to research pathways in Figure 5. Many species lack the large, well-characterised collections needed to train robust models, and rare species and emerging adulterants are particularly difficult (Peng et al., 2025; Tang et al., 2022); class imbalance degrades classifier performance and requires specialised resampling (Peng et al., 2025), while few-shot and transfer learning remain to be validated (Filho et al., 2023). The absence of public benchmark datasets hinders comparative evaluation and reproducibility (Liu et al., 2025; Wang et al., 2021), since most studies use proprietary collections with limited geographical and temporal representation (Liu et al., 2022; Kamal, 2022), and international reference collections with full metadata are needed (Durazzo et al., 2022; Lindenmaier et al., 2025).

Deep learning models function as black boxes, hindering mechanistic understanding and regulatory acceptance, since agencies require transparency to ensure reliability and prevent gaming (Durazzo et al., 2022; Lindenmaier et al., 2025; Vellido, 2020; Joshi, 2023). Models trained on one instrument often fail on another because of differences in spectral resolution, wavelength calibration and detector response, so transfer learning, domain adaptation and standardisation protocols are required (Wang and Yu, 2015). Most studies also use laboratory-grade instruments under controlled conditions, limiting applicability to supply chain work; translation to portable devices requires wavelength selection, model compression and validated robustness against temperature, humidity and vibration (Peng et al., 2025).

Preprocessing and feature selection choices significantly affect performance yet are often inadequately documented (Chen et al., 2023), and many studies rely on cross-validation alone without independent or external test sets, producing inflated estimates that do not generalise (Liu et al., 2025; Wang et al., 2021). Machine learning must also integrate with existing pharmacopoeial methods rather than replace them, and hybrid workflows combining rapid screening with confirmatory analysis offer a practical route (Figure 6) (Noviana et al., 2022; Klein-Junior et al., 2021). Finally, biological, seasonal and climate-driven variability challenges model stability and requires periodic retraining (Liu et al., 2022), while adulterators continue to develop schemes that evade classifiers trained on known adulterants, making one-class and anomaly detection necessary (Liu et al., 2021; Somov, 2025).

Table 4. Principal challenges, their consequences and recommended responses.

Challenge	Consequence	Recommended response
Small, imbalanced datasets	Poor generalisation; biased classifiers	Resampling within folds; transfer and few-shot learning

Challenge	Consequence	Recommended response
No standardised benchmarks	Comparisons across studies are not meaningful	International reference collections with full metadata
Limited interpretability	Regulatory rejection; no mechanistic insight	Attention, saliency and feature attribution reporting
Instrument variability	Models fail on a second instrument	Calibration transfer; domain adaptation
Laboratory-only validation	Performance not reproduced in the field	Portable-device validation under real conditions
Undocumented preprocessing	Results not reproducible	Report full preprocessing and feature selection pipeline
Cross-validation only	Inflated accuracy estimates	Independent and external test sets; prospective validation
Biological and temporal variability	Model drift across seasons and years	Periodic retraining; online and active learning
Evolving adulteration	Novel schemes evade trained classifiers	One-class and anomaly detection approaches

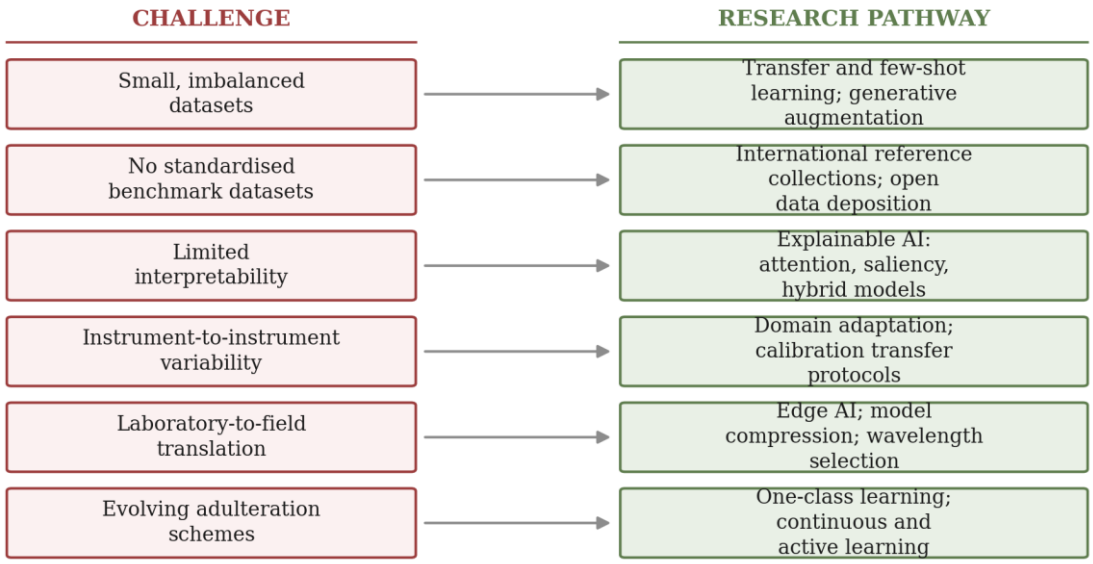

Figure 5. Mapping of the principal challenges identified in Section 8 to the corresponding research pathways proposed in Section 9.

9. FUTURE RESEARCH PATHWAYS

Explainable frameworks providing transparent interpretation of deep learning decisions are essential for regulatory acceptance (Vellido, 2020; Joshi, 2023). Attention mechanisms highlighting informative spectral regions, saliency maps and layer-wise relevance propagation offer practical routes (Kabir et al., 2022b), and hybrid models that combine mechanistic understanding with data-driven learning may further improve interpretability (Vellido, 2020). Transfer learning enables rapid adaptation to

new species with limited data (Filho et al., 2023; Tang et al., 2022), domain adaptation addresses instrument variability (Wang and Yu, 2015), and generative models can augment scarce training data (Filho et al., 2023).

Integration of metabolomics with proteomics, genomics and transcriptomics would capture genetic, biochemical and environmental influences together, requiring fusion strategies that accommodate different data types, scales and noise characteristics, with network-based approaches revealing systems-level quality determinants (Alum et al., 2025; Zhang and Wang, 2023). Blockchain enables immutable recording of quality data across the supply chain, and combining it with IoT monitoring of storage conditions provides end-to-end traceability with smart contracts automating acceptance criteria (Tichoniuk, 2024). International collaboration on reference collections, benchmark datasets and validation protocols, supported by regulatory guidance on acceptable methodologies, will facilitate adoption (Durazzo et al., 2022; Lindenmaier et al., 2025).

Edge computing enables deployment on resource-constrained portable devices, with pruning, quantisation and knowledge distillation reducing computational demand and wavelength selection simplifying instrument design (Peng et al., 2025). Practical workflows will use rapid screening to flag suspect samples for confirmatory chromatographic analysis, prioritised by uncertainty quantification and anomaly scores (Noviana et al., 2022; Somov, 2025), while online and active learning keep models current as supply chains evolve (Liu et al., 2022). Traditional knowledge systems offer quality indicators that can inform feature engineering, with natural language processing of classical texts extracting this knowledge and participatory approaches improving practical utility (Zhou et al., 2024a). Machine learning can finally support sustainable production by optimising cultivation, reducing waste and enabling traceability that rewards sustainable sourcing (Alum et al., 2025; Cazelles et al., 2020).

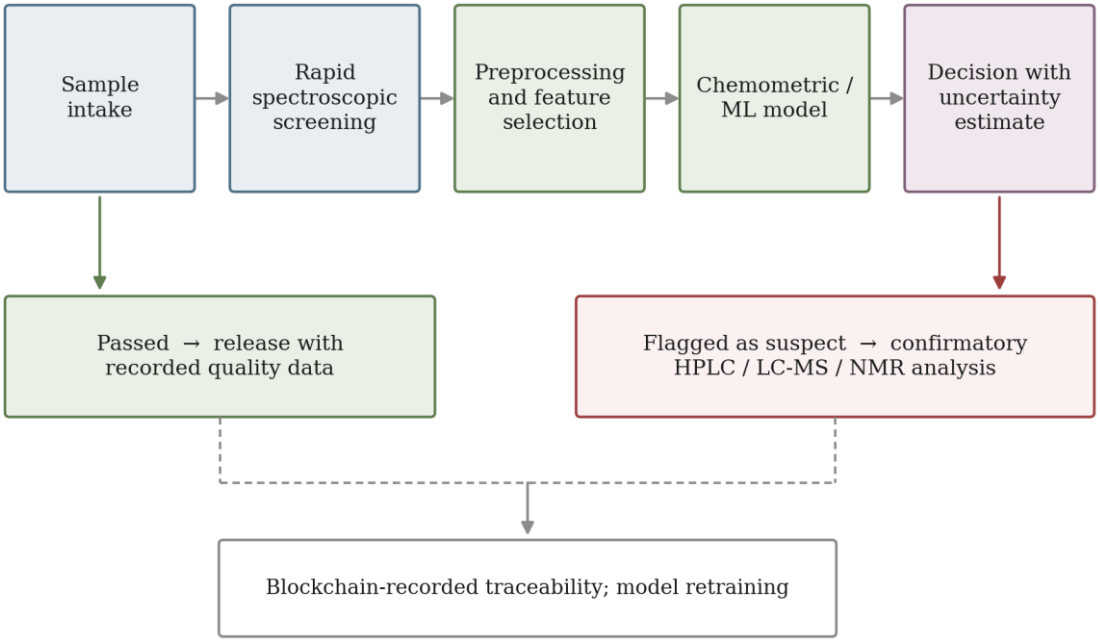

Figure 6. Proposed hybrid quality control workflow in which rapid spectroscopic screening triages samples and confirmatory chromatographic analysis is reserved for flagged material.

10. CONCLUSION

Machine learning and chemometrics matured into robust tools for herbal quality control between 2015 and 2025, outperforming traditional single-marker approaches across 150 reviewed papers. Vibrational spectroscopy with machine learning exceeded 95% accuracy for authentication and adulteration detection while remaining rapid and non-destructive; chromatographic and mass spectrometric methods enabled marker identification and metabolomics-driven standardisation; hyperspectral imaging approached 98–99% while adding spatial resolution; and data fusion improved discrimination, with mid-level fusion often optimal. Classical chemometrics retained value through interpretability and small-sample robustness, supervised learning improved accuracy, and deep learning achieved the highest reported performance.

Applications spanned authentication, adulteration detection, origin discrimination, processing quality, bioactive quantification and metabolomics-driven standardisation. Across all six, approaches leveraging the entire chemical fingerprint outperformed single-marker quantification, particularly for closely related species and subtle adulteration. Critical gaps nonetheless persist: small and imbalanced datasets, absent benchmark datasets, limited interpretability, instrument-to-instrument variability and unproven field translation.

Addressing these requires coordinated effort across explainable artificial intelligence, transfer learning, multi-omics integration, blockchain-enabled traceability, regulatory harmonisation and edge deployment. The convergence of advanced platforms, sophisticated computational methods and growing recognition of herbal complexity positions these technologies as transformative for quality control, provided the field converges on validation and reporting standards commensurate with the claims it makes.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the management of MIT Art, Design and Technology University Pune for providing the necessary support and resources to successfully complete this research.

AUTHOR CONTRIBUTIONS

Author One: Conceptualization, Literature Search, Formal Analysis, Author two: Writing — Original Draft, Visualization. Author Three: Methodology, Literature Screening, Author Four, Five and Six: Writing — Review & Editing, Supervision. All authors have read and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

This review did not involve human participants, animal subjects or the collection of primary data, and ethical approval was therefore not required.

DATA AVAILABILITY STATEMENT

No new data were generated for this review. All studies discussed are available from their original publishers, and the search strategy is described in Section 3.

REFERENCES

[1] Abraham, E.J., Kellogg, J.J., 2021. Chemometric-guided approaches for profiling and authenticating botanical materials. Frontiers in Nutrition 8, 780228.

[2] Alum, E.U., Manjula, V., Uti, D.E., et al., 2025. Metabolomics-driven standardization of herbal medicine: advances, applications, and sustainability considerations. Natural Product Communications 20, 1934578X251367650.

[3] Arifah, M.F., Hastuti, A.A.M.B., Rohman, A., 2022. Application of molecular spectroscopy and chemometrics for authentication and quality control of tea (*Camellia sinensis* L.): a review. Indonesian Journal of Pharmacy 33(2), doi:10.22146/ijp.4817.

[4] Bai, L., Zhang, Z.-T., Guan, H., et al., 2024. Rapid and accurate quality evaluation of Angelicae Sinensis Radix based on near-infrared spectroscopy and Bayesian optimized LSTM network. Talanta 275, 126098.

[5] Bai, Y., Zhang, H., 2024. The cluster analysis of traditional Chinese medicine authenticity identification technique assisted by chemometrics. Heliyon 10, e37479.

[6] Calixto, M.G., Ramos, H.A., Veríssimo, L.S., et al., 2021. Trends and application of chemometric pattern recognition techniques in medicinal plants analysis. Critical Reviews in Analytical Chemistry, doi:10.1080/10408347.2021.1953370.

[7] Camps, C., 2020a. FT-NIR assisted machine and deep learning for determination of acteoside, aucubin and catalpol contents of *Plantago lanceolata*, Agricultural Research and Technology 25, doi:10.19080/artoaj.2020.25.556296.

[8] Cazelles, K., Zemlak, T.S., Gutgesell, M., et al., 2020. Spatial fingerprinting: horizontal fusion of multi-dimensional bio-tracers as solution to global food provenance problems. bioRxiv, doi:10.1101/2020.10.01.322453.

[9] Chen, F., Zhang, M., Huang, W., et al., 2024. Laser-induced breakdown spectroscopy–visible and near-infrared spectroscopy fusion based on deep learning network for identification of adulterated Polygonati Rhizoma. Foods 13(14), 2306.

[10] Chen, R., Liu, M., Zhang, C., et al., 2023. Trends in digital detection for the quality and safety of herbs using infrared and Raman spectroscopy. Frontiers in Plant Science 14, 1128300.

[11] Choi, J.-Y., Lim, J., Yang, J.-H., et al., 2025. Hyperspectral imaging combined with machine and deep learning for non-destructive detection of defective red pepper (*Capsicum annuum* L.) powder adulteration. International Journal of Food Properties, doi:10.1080/10942912.2025.2564386.

[12] Ding, R., Wang, C., Zhong, S.-H., et al., 2023. Quality assessment of traditional Chinese medicine based on data fusion combined with machine learning: a review. Critical Reviews in Analytical Chemistry 54(5), doi:10.1080/10408347.2023.2189477.

[13] Durazzo, A., Sorkin, B.C., Lucarini, M., et al., 2022. Analytical challenges and metrological approaches to ensuring dietary supplement quality: international perspectives. Frontiers in Pharmacology 12, 714434.

[14] Fernandes, J.M., Silvestre, C., Zucolotto, S.M., et al., 2025. Flavonoids as markers in herbal medicine quality control: current trends and analytical perspective. Separations 12(11), 289.

[15] Filho, A.R.G., Provensi, L.R., Rosseto, L.P., et al., 2023. Quality control based on species classification of herbal medicines using hyperspectral imaging and machine learning. In: Proceedings of IEEE WHISPERS, doi:10.1109/whispers61460.2023.10431013.

[16] Gan, Y., Yang, T., Gu, W., et al., 2024. Using HS-GC-MS and flash GC e-nose in combination with chemometric analysis and machine learning algorithms to identify the varieties, geographical origins and production modes of *Atractylodes lancea*. Industrial Crops and Products 207, 117955.

[17] Guan, H., Zhang, Z.-T., Bai, L., et al., 2024. Multi-spectra combined with Bayesian optimized machine learning algorithms for rapid and non-destructive detection of adulterated functional food *Panax notoginseng* powder. Journal of Food Composition and Analysis 133, 106412.

[18] Hao, J.-W., Liu, X.-Q., Liu, X.-Q., et al., 2021. Rapid detection of adulteration in *Dendrobium* huoshanense with *Dendrobium* henanense by ATR-FTIR combined with multivariate methods. Iranian Journal of Pharmaceutical Research 20(3), 291–303.

[19] He, M., Li, S.-P., 2024. Chemometrics in quality control of traditional Chinese medicines. In: Medicinal Plants: Chemistry, Biology and Omics. Springer, doi:10.1007/978-981-99-9871-5_20.

[20] Hu, S., Li, H., Chen, C., et al., 2022. Raman spectroscopy combined with machine learning algorithms to detect adulterated Suichang native honey. Scientific Reports 12, 3456.

[21] Jaiswal, A., Bhardwaj, S., Kandasamy, Nagarajan, N., 2025. AI/ML-based quality control testing for standardization of herbals. Current Traditional Medicine, doi:10.2174/0129504910376010251012115617.

[22] Ji, J., Meng, J., Zhong, J., et al., 2024. Discrimination method of processed traditional Chinese medicine products based on near-infrared spectroscopy and deep learning. In: Proceedings of IEEE ICNC-FSKD, doi:10.1109/icnc-fskd64080.2024.10702262.

[23] Joshi, P.B., 2023. Navigating with chemometrics and machine learning in chemistry. Artificial Intelligence Review 56, 6985–7034.

[24] Kabir, M., Guindo, M.L., Chen, R., et al., 2022a. Application of laser-induced breakdown spectroscopy and chemometrics for the quality evaluation of foods with medicinal properties: a review. Foods 11(14), 2051.

[25] Kabir, M., Guindo, M.L., Chen, R., et al., 2022b. Deep learning combined with hyperspectral imaging technology for variety discrimination of *Fritillaria thunbergii*. Molecules 27(18), 6042.

[26] Kamal, N., 2022. Current authentication methods of herbs and herbal products: a systematic review. Food Research 6(4), doi:10.26656/fr.2017.6(4).468.

[27] Kaur, T., Kaur, S., Khan, W.A., et al., 2026. Analytical techniques for quality control of botanical extracts. In: Quality Control in Herbal Medicine. CRC Press, doi:10.1201/9781003662785-6.

[28] Klein-Junior, L.C., de Souza, M.R., Viaene, J., et al., 2021. Quality control of herbal medicines: from traditional techniques to state-of-the-art approaches. Planta Medica 87(11), 803–826.

[29] Kusumadewi, A., Martien, R.F., Pramono, S., et al., 2022. Review on the application of chemometrics for the standardization and authentication of *Curcuma xanthorrhiza*. Food Research 6(2), doi:10.26656/fr.2017.6(2).044.

[30] Lam, W.-C., Cai, W., Li, Y., et al., 2023. Editorial: quality control for efficacy and safety of herbal medicinal products. Frontiers in Pharmacology 14, 1162698.
[31] Liang, Q., Liu, S., 2025. Research progress in AI-assisted characteristic pattern analysis and quality evaluation methods for authentic medicinal materials. Innovative Applications of AI, doi:10.70695/iaai202503a7.
[32] Lindenmaier, M., Bernart, M.W., Brinckmann, J.A., 2025. Advanced methodologies for the quality control of herbal supplements and regulatory considerations. Phytochemical Analysis, doi:10.1002/pca.70000.
[33] Liu, C., Zuo, Z.-T., Xu, F., et al., 2022. Authentication of herbal medicines based on modern analytical technology combined with chemometrics approach: a review. Critical Reviews in Analytical Chemistry 53(4), doi:10.1080/10408347.2021.2023460.
[34] Liu, X., Liu, X., Wang, J., et al., 2025. Machine learning and chemometric methods for high-throughput authentication of 53 root and rhizome Chinese herbal using ATR-FTIR fingerprints. Journal of Chromatography B, doi:10.1016/j.jchromb.2025.124630.
[35] Liu, Z., Yang, M.Q., Zuo, Y., et al., 2021. Fraud detection of herbal medicines based on modern analytical technologies combined with chemometrics approach: a review. Critical Reviews in Analytical Chemistry 52(5), doi:10.1080/10408347.2021.1905503.
[36] Mall, P., Chamoli, V., Raina, D., 2025. Artificial intelligence-powered precision: predicting standard curves for herbal medicines. Social Science Research Network, doi:10.2139/ssrn.5083338.
[37] Noviana, E., Indrayanto, G., Rohman, A., 2022. Advances in fingerprint analysis for standardization and quality control of herbal medicines. Frontiers in Pharmacology 13, 853023.
[38] Peng, C., Zhang, M., Kong, M., et al., 2024. Rapid quantitative analysis of active ingredients in Radix Gentianae Macrophyllae extracts based on near-infrared spectroscopy and deep learning. SSRN, doi:10.2139/ssrn.4820199.
[39] Peng, X., Yu, X., Lin, J., et al., 2025. Rapid authentication of Cuscutae Semen with imbalanced sample categories using a handheld near infrared spectrometer. Journal of Near Infrared Spectroscopy, doi:10.1177/09670335251314161.
[40] Rebiai, A., Seghir, B.B., Hemmami, H., et al., 2022. Quality assessment of medicinal plants via chemometric exploration of quantitative NMR data: a review. Compounds 2(2), 128–145.
[41] Ren, L., Wang, F., Zhang, Y., et al., 2024. Rapid identification of Radix Astragali by data fusion of laser-induced breakdown spectroscopy and Raman spectroscopy coupled with deep learning. Talanta 280, 127016.
[42] Riswanto, F.D.O., Windarsih, A., Lukitaningsih, E., et al., 2022. Metabolite fingerprinting based on 1H-NMR spectroscopy and liquid chromatography for the authentication of herbal products. Molecules 27(4), 1198.
[43] Singh, A., Agarwal, K., Kumar, A., et al., 2025. Advanced methods for standardizing herbal medicines: a detailed review. International Journal for Multidisciplinary Research 7(6), doi:10.36948/ijfmr.2025.v07i06.63231.
[44] Somov, D.V., 2025. A one-class autoencoder neural network approach for rapid Raman-based pharmaceutical quality screening. Journal of Raman Spectroscopy, doi:10.1002/jrs.70065.
[45] Taha, H., Abu-Khalaf, N., 2020. Quality control for herbal medicinal plants using a sensor array (an electronic tongue). In: Proceedings of ICBE, doi:10.46617/icbe7001.
[46] Tang, J.-W., Hong, Y.-X., Chen, J., et al., 2025. Ratio quantification of geranium and rose essential oil mixtures via deep learning analysis of complex Raman spectra. ACS Measurement Science Au, doi:10.1021/acsmeasuresciau.5c00088.
[47] Tang, S., Jue, J., Lin, Y., 2022. Study on identification method of Chinese herbal medicine based on infrared spectroscopy characteristics. Highlights in Science, Engineering and Technology 24, doi:10.54097/hset.v24i.3906.
[48] Teklemariam, T.A., 2024. Raman and mid-infrared spectroscopy coupled with machine–deep learning for adulterant detection in ground turmeric. Journal of Vibrational Spectroscopy, doi:10.1177/27551857241250014.
[49] Tichoniuk, M., 2024. Digital product passport (DPP) as an important mechanism supporting the circular economy. doi:10.56091/ctqs.innov-21.
[50] Vellido, A., 2020. The importance of interpretability and visualization in machine learning for applications in medicine and health care. Neural Computing and Applications 32, 18069–18083.
[51] Wang, H., Chen, Y., Wang, L., et al., 2023. Advancing herbal medicine: enhancing product quality and safety through robust quality control practices. Frontiers in Pharmacology 14, 1265178.
[52] Wang, P., Yu, Z., 2015. Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: a review. Journal of Pharmaceutical Analysis 5(5), 277–284.

[53] Wang, Y., Zuo, Z.-T., Wang, Y., 2021. Pattern recognition: an effective tool for quality assessment of herbal medicine based on chemical information. Journal of Chemometrics 35(3), e3305.

[54] Wu, H.-Y., Shaw, P.-C., 2022. Strategies for molecular authentication of herbal products: from experimental design to data analysis. Chinese Medicine 17(1), 32.

[55] Yang, L., Xue, Y., Wei, J., et al., 2021. Integrating metabolomic data with machine learning approach for discovery of Q-markers from Jinqi Jiangtang preparation against type 2 diabetes. Chin. Med. 16(1), article 30 (2021), DOI 10.1186/s13020-021-00438-x.

[56] Yang, P., Zeng, Z., Hou, Y., et al., 2023. Rapid authentication of variants of *Gastrodia elata* Blume using near-infrared spectroscopy combined with chemometric methods. Journal of Pharmaceutical and Biomedical Analysis 234, 115592.

[57] Zhai, D., Zhou, J., Ma, J., et al., 2025. Rapid identification of adulteration in American ginseng powder using near-infrared spectroscopy combined with machine learning. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 337, 127409.

[58] Zhan, G.-Z., Guo, X.-Y., Qiu, Z.-C., et al., 2024. Rapid evaluation of the quality of *Smilax glabra* Roxb. using QADS based on FT-NIR combined with multiple intelligent algorithms. Food Chemistry 444, 139633.

[59] Zhang, Y., Wang, Y., 2023. Recent trends of machine learning applied to multi-source data of medicinal plants. Journal of Pharmaceutical Analysis 13(12), 1388–1407.

[60] Alkhatib A. J., (2026). Integrated Computational Retinal Vascular Morphometry Using Fractal, Skeleton-Based, and Phi-Related Features for Diabetic Retinopathy and Glaucoma. International Journal of Artificial Intelligence, Machine Learning and Data Analytics, 1(1), 43-50. https://www.ijaimda.org/index.php/ijaimda/article/view/3

[61] Zhao, Y., Qian, Z., Li, X., et al., 2025. Rapid authentication of *Dendrobium* species using attenuated total reflectance Fourier transform infrared (ATR-FTIR) and Fourier transform near-infrared (FT-NIR) spectroscopies with orthogonal partial least squares–discriminant analysis (OPLS-DA). figshare, doi:10.6084/m9.figshare.28294695.

[62] Zhou, E., Shen, Q., Hou, Y., et al., 2024a. Integrating artificial intelligence into the modernization of traditional Chinese medicine industry: a review. Frontiers in Pharmacology 15, 1181183.

[63] 63. Singh. R., Verghese. S. P., Bansal A. (2026). Silver Nanoparticles For Environmental Sustainability And Their Role In Biological Activities. International Journal of Engineering Sciences & Research Technology, 15(3), 1–8. ttps://doi.org/10.64149/j.ijesrt.15.3-1-8

[64] Zhou, F., Xie, W., Lin, M.-H., et al., 2024b. Rapid authentication of geographical origins of Baishao (Radix Paeoniae Alba) slices with laser-induced breakdown spectroscopy based on conventional machine learning and deep learning. Spectrochimica Acta Part B 212, 106852.

[65] Zhu, S., Liu, Q., Qiu, S., et al., 2022. DNA barcoding: an efficient technology to authenticate plant species of traditional Chinese medicine and recent advances. Chinese Medicine 17(1), 111.