

Machine Learning-Based Quality Evaluation and Authenticity Identification of Traditional Chinese Medicinal Materials

Manish J. Malik

School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.

manish.work@uc.edu

Samuel Dawson

Department of Computer Science, University of North Texas, Denton, TX, USA.

sdawson@unt.edu

Emmett Sorensen

Department of Computer Science, University of Central Florida, Orlando, FL, USA.

emmetts@ucf.edu

Abstract

The authentication and quality evaluation of traditional Chinese medicinal materials remain deeply challenging because these materials exhibit complex biological, chemical, and geographic variability, while commercial supply chains are vulnerable to substitution, adulteration, and mislabeling. Machine learning offers a systematic pathway to integrate heterogeneous analytical signals derived from spectroscopy, chromatography, imaging, genomic markers, and chemical fingerprinting. However, real-world deployment demands attention to far more than predictive accuracy. This paper presents a system-level examination of machine learning architectures for medicinal material quality evaluation and authenticity identification. It discusses data infrastructure, multi-source sensing, representation learning, interpretability, deployment trade-offs, governance, fairness, and long-term sustainability. The analysis emphasizes that robust systems must balance high-dimensional model capacity against operational constraints such as limited reference samples, evolving adulteration patterns, regulatory traceability, and cross-regional fairness. Deep neural networks, ensemble classifiers, sequence models, and attention-based architectures all provide distinct advantages depending on the sensing modality and reliability requirements. At the same time, interpretability tools and data provenance mechanisms are essential for regulatory acceptance and scientific credibility. The paper therefore foregrounds socio-technical integration rather than isolated algorithmic performance. By treating quality evaluation as a data-intensive infrastructure problem, the discussion offers architectural guidance for researchers, standards bodies, pharmaceutical regulators, and supply chain stakeholders seeking to build trustworthy machine learning systems for traditional medicine.

Keywords

traditional Chinese medicine; machine learning; authenticity identification; quality evaluation; data infrastructure; system governance.

1. Introduction

Traditional Chinese medicinal materials occupy a distinctive position in global pharmaceutical and wellness markets because they are simultaneously biological products, cultural artifacts, and complex chemical mixtures. Quality evaluation has historically depended on organoleptic inspection, morphological observation, microscopic examination, and chromatographic fingerprinting. These methods remain valuable but are often subjective, labor-intensive, and difficult to standardize across large volumes of material. The globalization of herbal supply chains has amplified the need for automated, scalable, and evidence-driven systems capable of detecting adulterants, distinguishing closely related species, and estimating active constituent profiles. Recent advances in representation learning and deep neural architectures have opened new possibilities for learning hierarchical features from complex instrumental inputs [1]. Such methods reduce the reliance on manually engineered descriptors and allow models to capture subtle patterns in high-dimensional data [2]. However, traditional machine learning methods still retain significant practical value in this domain. Ensemble classifiers such as random forests can deliver robust performance across heterogeneous tabular chemical profiles, while support vector networks remain relevant for small-sample spectroscopic discrimination tasks [3,4]. Additional analytical approaches, including the use of cell wall compositional profiles, have extended the set of phenotypic markers available for distinguishing closely related medicinal plants [5]. The present paper therefore approaches machine learning for medicinal material quality evaluation as a systems problem rather than a narrow classification task.

The central argument is that effective deployment requires coordinated design across data acquisition, model selection, validation, interpretability, and governance. A model that performs well on a curated spectral library may fail in field settings because source materials vary by harvest season, processing method, storage condition, and geographic origin. Furthermore, authentication is not simply a technical classification problem. It is also a regulatory and economic problem because incorrect identity or quality labels can lead to patient harm, loss of market trust, and legal liability. The machine learning community has produced powerful tools for image classification, sequence modeling, and tabular prediction, but translating these tools into traditional medicine systems requires careful attention to data scarcity, domain shift, and scientific interpretability. This paper examines structural trade-offs across sensing architectures, model families, deployment environments, and regulatory frameworks. By integrating perspectives from analytical chemistry, pharmacognosy, machine learning, and socio-technical infrastructure studies, the discussion provides a publication-oriented synthesis of current capabilities and future directions.

2. From Pharmacognosy to Data-Driven Quality Assessment

Pharmacognosy has long relied on the principle that botanical identity and medicinal quality are encoded in observable structural and chemical features. Macroscopic and microscopic traits, thin-layer chromatographic patterns, and high-performance liquid chromatographic profiles all provide useful information about species identity and constituent content. Yet these conventional approaches encounter structural limits when faced with complex multi-herb formulations, processed slices, powders, extracts, and materials whose morphological features have been partially destroyed. In such cases, chemical profiling becomes central. Chemometric analysis of chromatographic and spectroscopic data has been used to classify medicinal species and detect adulteration, but its effectiveness depends on the stability of reference libraries and the quality of preprocessing. Gradient-boosted tree ensembles have also gained attention for high-dimensional chemical fingerprint classification because they

combine strong predictive performance with variable importance measures that can support chemical interpretation [6]. These models are especially useful when the data are tabular, heterogeneous, and contain missing values arising from inconsistent analytical runs.

The transition toward data-driven quality assessment does not eliminate the need for domain knowledge. Instead, it transforms domain knowledge into data curation strategies, feature definitions, and validation protocols. A machine learning system intended to distinguish authentic materials from substitutes must encode assumptions about what constitutes representative variation. If the training set contains only a narrow subset of geographic origins or processing methods, the resulting model may mistake unfamiliar authentic samples for adulterants. Conversely, if the training set is too broad and includes poorly characterized commercial samples, the model may learn patterns associated with packaging or storage rather than intrinsic material quality. Therefore, the design of reference data collections is as important as the choice of algorithm. Data provenance, analytical instrument metadata, extraction conditions, and sampling location should be recorded alongside spectral or chemical measurements. These infrastructural considerations are often undervalued in proof-of-concept studies yet become decisive in operational systems.

3. Data Infrastructure and Multi-Source Sensing Architectures

A practical machine learning system for medicinal material authentication must accommodate multiple sensing modalities. Imaging systems can capture macroscopic surface features such as texture, color, and shape, while microscopic imaging reveals tissue structures and diagnostic cell types. Spectroscopic platforms provide chemical information at different molecular levels, including near-infrared, mid-infrared, Raman, and nuclear magnetic resonance signatures. Chromatographic methods produce time-resolved peaks that reflect constituent composition. Genomic and metabolomic assays add further layers of biological identity. The challenge lies in fusing these modalities without creating brittle dependencies. Deep convolutional networks have demonstrated strong performance in image-based classification tasks and can be adapted to macroscopic or microscopic images of medicinal materials [7]. These architectures learn spatially hierarchical features that may correspond to tissue structures, surface patterns, or powder particle distributions [8]. When trained on sufficiently large and diverse image collections, they can achieve high accuracy even when visual differences are subtle.

Beyond image data, modern sequence-oriented architectures such as transformers have transformed the analysis of structured and sequential inputs [9]. In the medicinal materials context, transformers can be applied to spectral sequences, chromatographic time series, and even structured textual records describing botanical origin, processing method, and historical authentication results. Pretrained models originally developed for language have inspired domain-specific pretraining strategies that learn contextual representations from large corpora of analytical records [10]. These representations can then be fine-tuned for downstream tasks such as species classification, adulteration detection, or quality grade prediction. The advantage of such approaches is their ability to model long-range dependencies and to transfer knowledge from unlabeled data. However, they require substantial computational infrastructure and careful management of data heterogeneity.

Interpretability layers should be integrated into the data infrastructure rather than treated as post hoc accessories. For regulatory acceptance, it is insufficient to report a predicted label without explaining which spectral regions, image patches, or chemical peaks contributed most to the decision. Attribution methods such as SHAP provide additive feature importance

estimates for complex models [11]. Local interpretable model-agnostic explanations can generate simplified local approximations that help analysts understand individual predictions [12]. Gradient-based localization techniques can highlight discriminative image regions in convolutional networks [13]. These methods allow pharmacognosists and quality control scientists to evaluate whether a machine learning model is attending to scientifically meaningful patterns or to incidental artifacts. When interpretability outputs are logged as part of a quality management system, they support auditability and continuous improvement.

4. Machine Learning Architectures for Quality and Authentication

The selection of machine learning architectures should be guided by the temporal, spatial, and chemical structure of the available data. For time-resolved sensor signals and chromatographic profiles, recurrent networks such as long short-term memory models have historically provided strong performance because they can retain information across variable-length sequences [14]. These models can be used to align chromatographic peaks, detect anomalies in elution patterns, and classify samples based on dynamic response trajectories. At the same time, chemometric approaches continue to provide interpretable baselines. Reviews of chemometrics in herbal authentication have emphasized the importance of data preprocessing, sample representativeness, and validation strategy [15]. Chemical markers remain central to quality control because they connect classification outputs to known pharmacologically active constituents and can be verified using independent analytical methods [16]. Machine learning models that integrate both marker-based chemical features and raw spectral representations may therefore achieve a better balance between interpretability and predictive performance.

Modern convolutional architectures supply additional capacity for spectral imaging and high-resolution chemical mapping. Very deep convolutional networks have shown that increased depth can improve representational power when sufficient data are available [17]. Multi-scale convolutional designs can capture both fine-grained local structures and broader spatial contexts, which is particularly relevant for images that contain mixed botanical fragments or heterogeneous powder distributions [18]. These models can be used to classify species from high-resolution images, detect foreign matter, and estimate visual quality grades. Nonetheless, their deployment requires careful attention to calibration against reference standards and to the risk of learning spurious correlations from background conditions, lighting, or acquisition equipment. Domain adaptation and regularized retraining are therefore essential components of any long-term deployment architecture.

5. System-Level Trade-Offs and Deployment

Deploying machine learning systems in traditional medicine supply chains involves multiple structural trade-offs. Model capacity must be weighed against the availability of labeled reference data. Deep architectures provide flexibility but demand large training sets, whereas simpler models may be more stable when authentic samples are scarce. Latency requirements also differ across use cases. A field inspector may require near real-time screening on a portable device, while a centralized regulatory laboratory can tolerate more computationally intensive batch analysis. Dimensionality reduction techniques such as t-distributed stochastic neighbor embedding can support monitoring of data drift by projecting learned representations into two-dimensional spaces where domain experts can visually inspect clustering and outlier patterns [19]. Drift monitoring is especially important because seasonal variation, new adulterants, and changes in processing methods can gradually shift the data distribution away from that used during initial training.

Software infrastructure choices influence reproducibility, scalability, and security. Deep learning frameworks such as TensorFlow and PyTorch have become widely adopted because they support flexible model development, distributed training, and deployment across heterogeneous hardware [20,21]. These ecosystems also provide model serialization formats, serving tools, and integration interfaces that simplify the transition from research prototypes to production systems. However, operational systems must include additional layers for data validation, model versioning, audit logging, and access control. Without these layers, a technically accurate model may still produce unreliable decisions because input data are corrupted, preprocessing steps are changed, or unauthorized personnel modify the inference pipeline. A system-level design therefore treats machine learning as one component within a larger quality assurance workflow that includes standard operating procedures, instrumentation calibration, and personnel training.

6. Governance, Fairness and Sustainability

Governance is a decisive factor in the adoption of machine learning for medicinal material authentication. Regulatory frameworks for herbal substances and herbal medicinal products establish expectations for identity testing, purity, and constituent specification [22]. Machine learning outputs must be aligned with these frameworks through validation reports, reference standards, and documented uncertainty. Fairness is also a critical but often overlooked concern. If training data are dominated by materials from a small number of regions or manufacturers, the resulting models may systematically disadvantage underrepresented producers whose authentic materials exhibit legitimate variation. This can create economic barriers and reduce confidence in the system. Fairness-aware evaluation should therefore assess model performance across geographic origins, production scales, and processing methods. Governance mechanisms should also address data ownership, benefit sharing, and the protection of traditional knowledge associated with medicinal plants.

Sustainability requires thinking beyond immediate prediction accuracy. The construction of high-quality reference libraries is resource intensive, and the resulting data assets must be maintained over time. Instruments change, calibration protocols evolve, and new adulteration strategies emerge. A sustainable system therefore includes mechanisms for periodic retraining, expert review of uncertain cases, and open publication of validation results where appropriate. Cloud-based infrastructures may offer scalability, but they also introduce dependencies on external providers and energy consumption concerns. Edge computing approaches can reduce data transfer and improve responsiveness, but they may limit model complexity and complicate model updates. A balanced deployment may use lightweight screening models at the point of inspection and reserve more sophisticated multimodal models for confirmatory analysis in centralized laboratories. These architectural decisions should be guided by risk assessment, cost constraints, and the clinical or commercial consequences of misclassification.

7. Conclusion

Machine learning has the potential to transform the quality evaluation and authenticity identification of traditional Chinese medicinal materials by integrating heterogeneous analytical data into coherent decision systems. However, this potential can only be realized when algorithmic development is embedded within a broader socio-technical infrastructure. The preceding analysis has emphasized that data provenance, sensing architecture, model interpretability, deployment constraints, governance, fairness, and sustainability are not peripheral concerns but central design variables. Deep convolutional networks, recurrent models, gradient-boosted ensembles, and attention-based architectures each offer distinct

capabilities, but no single model family is universally superior. The appropriate architecture depends on the modality, data volume, interpretability requirements, and operational environment. By focusing on systems rather than isolated classifiers, researchers and practitioners can build machine learning solutions that are scientifically credible, regulatory compatible, and operationally durable. Future work should prioritize the construction of well-annotated multi-source reference libraries, the development of standardized validation protocols, and the integration of machine learning outputs into transparent quality management systems.

References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
3. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
4. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
5. Zhao, K., Li, G., Li, C., Jiang, T., Wang, W., & Yang, Z. (2026). Identification Markers for *Salvia miltiorrhiza* and Its Close Relatives Based on Cell Wall Component Characteristics. *Engineered Science*, 40, 2122.
6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. <https://doi.org/10.1145/2939672.2939785>
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778). IEEE. <https://doi.org/10.1109/CVPR.2016.90>
8. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25* (pp. 1097-1105). Curran Associates.
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30* (pp. 5998-6008). Curran Associates.
10. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171-4186). Association for Computational Linguistics.
11. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765-4774). Curran Associates.
12. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). ACM. <https://doi.org/10.1145/2939672.2939778>

13. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 618-626). IEEE. <https://doi.org/10.1109/ICCV.2017.74>
14. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
15. Gad, H. A., El-Ahmady, S. H., Abou-Shoer, M. I., & Al-Azizi, M. M. (2013). Application of chemometrics in authentication of herbal medicines: A review. *Phytochemical Analysis*, 24(1), 1-24. <https://doi.org/10.1002/pca.2378>
16. Li, S., Han, Q., Qiao, C., Song, J., Lunga, C. K., & Xu, H. (2008). Chemical markers for the quality control of herbal medicines: An overview. *Chinese Medicine*, 3, Article 7. <https://doi.org/10.1186/1749-8546-3-7>
17. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In 3rd International Conference on Learning Representations. ICLR.
18. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2818-2826). IEEE. <https://doi.org/10.1109/CVPR.2016.308>
19. Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579-2605.
20. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.
21. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems 32 (pp. 8024-8035). Curran Associates.
22. European Medicines Agency. (2011). Guideline on specifications: Test procedures and acceptance criteria for herbal substances, herbal preparations and herbal medicinal products/traditional herbal medicinal products. European Medicines Agency.