

Authenticity assessment of botanical materials using One-Class Support Vector Machine Classification in *visionCATS*

Example: Detection of non-conforming red Asian ginseng

Keywords

HPTLC, botanical authentication, fingerprint analysis, machine learning, one-class classification, one-class support vector machine (OC-SVM), quality control, red Asian ginseng

Introduction

The authentication of botanical materials is an important aspect of quality control in the herbal and nutraceutical industries. High-value botanicals may be particularly susceptible to substitution, adulteration, or misidentification, while their chemical composition can also vary as a result of processing, geographical origin, cultivation conditions, or storage. Chromatographic fingerprinting is widely used to assess the identity and chemical consistency of botanical materials by comparing characteristic profiles between samples. Although visual comparison by experienced analysts can be effective, interpretation becomes increasingly challenging when large numbers of samples are evaluated or when differences between chromatographic fingerprints are subtle.

Chemometric and machine-learning approaches provide a standardized and data-driven framework for evaluating chromatographic fingerprints. Among these approaches, One-Class Classification (OCC) is particularly well suited for authenticity assessment because comprehensive collections of all possible adulterants and non-conforming materials are rarely available. Rather than learning the characteristics of multiple classes, OCC models the distribution of an authenticated reference population and evaluates whether the fingerprint of an unknown sample is consistent with this reference class.

This application note demonstrates the implementation of the One-Class Support Vector Machine (OC-SVM) algorithm available in CAMAG *visionCATS* for the authenticity assessment of red Asian ginseng, *Panax ginseng*, using HPTLC fingerprint data. The approach establishes a decision model based exclusively on authenticated reference samples and subsequently evaluates unknown samples against the acceptance region defined by this reference set. Samples falling outside this acceptance region are flagged as non-conforming.

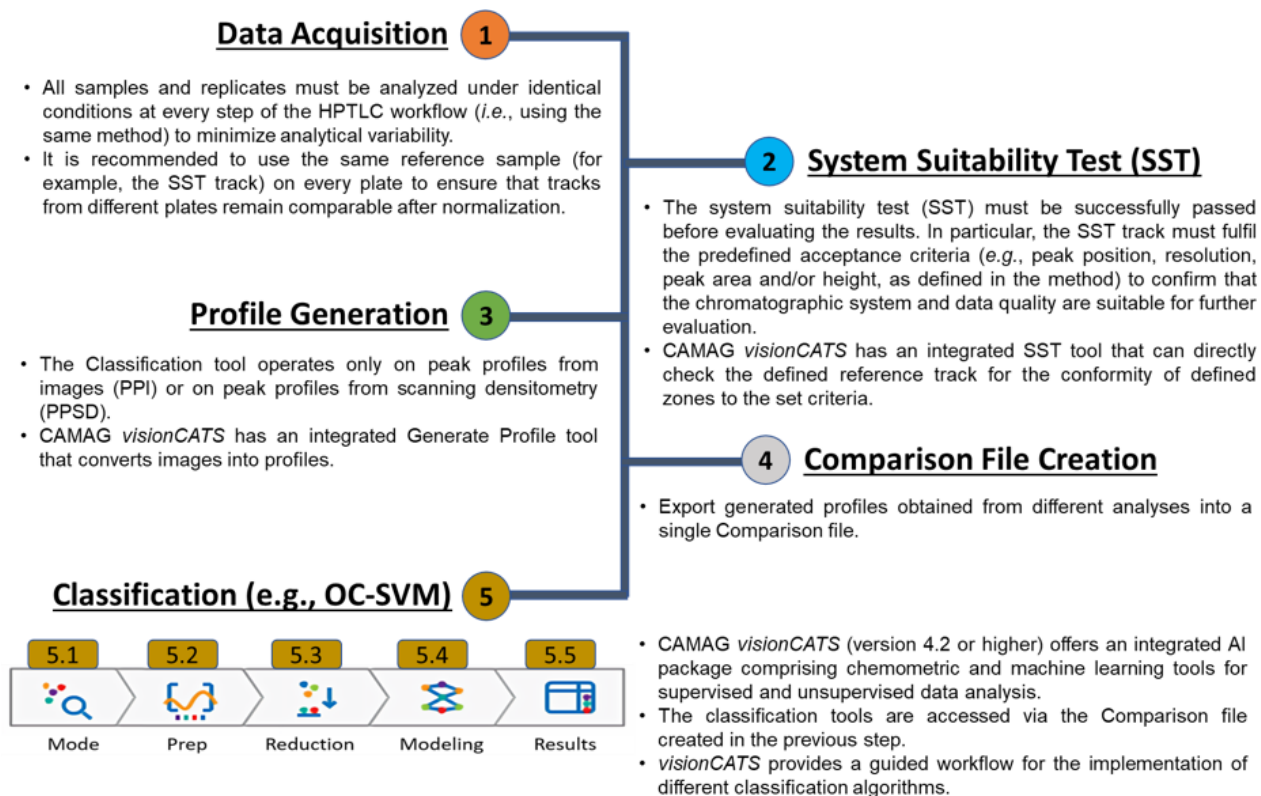
The presented workflow illustrates how HPTLC fingerprinting and machine-learning-based classification can be combined to provide a standardized and data-driven approach for authenticity assessment, even when representative examples of all possible adulterants are unavailable.

Scope

This application note demonstrates the use of the One-Class Support Vector Machine (OC-SVM) for the authenticity assessment of red Asian ginseng. The workflow illustrates how authenticated HPTLC fingerprint profiles can be used to establish an acceptance region and to identify unknown samples whose fingerprints fall outside the variability represented by the reference set.

Authenticity Assessment Workflow

A structured workflow consisting of data acquisition, system suitability verification, profile generation, comparison file creation, data preprocessing, OC-SVM model training, and classification of unknown samples was followed.



visionCATS workflow used for OC-SVM classification. The classification process comprises selection of the analysis mode (5.1), data preprocessing (5.2), optional dimensionality reduction (5.3) (not applied in this study), model definition and training (5.4), and evaluation of classification results (5.5).

1. Data Acquisition

Required or recommended devices

Automatic TLC Sampler 4, Automatic Developing Chamber ADC 2, Derivatizer, TLC Plate Heater 3, TLC Visualizer 2, *visionCATS* 4.2 or higher.

Preparation of samples

Mix 1.0 g of finely powdered sample with 5.0 mL of absolute ethanol and vortex for 30 s. Sonicate for 15 minutes and subsequently centrifuge the extract. The resulting supernatant is used as the test solution.

Preparation of standard

System Suitability Test (SST): Universal HPTLC Mix (UHM) prepared in-house (also commercially available from Sigma-Aldrich, cat. no. 91816).

Chromatography

Stationary phase	HPTLC Si 60 F ₂₅₄ , 20 × 10 cm (Supelco)
Sample application	8.0 mm bands with ATS 4, 15 tracks, track distance 11.4 mm, distance from left edge 20.0 mm, distance from lower edge 8.0 mm, application volume of 5.0 µL of the sample solution and 2.0 µL of the SST solution.
Developing solvent	Methanol, propyl acetate, tetrahydrofuran, water 2.5:15:15:3 (V/V/V/V).

Development	Develop in the ADC 2 after 20 minutes of chamber saturation using a saturation pad, without plate activation.
Developing distance	70.0 mm (from the lower edge)
Plate drying	Drying for 5 min in the ADC 2
Derivatization	10% Sulfuric acid reagent in methanol Preparation: Carefully mix 90.0 mL of cold methanol with 10.0 mL of concentrated sulfuric acid. Application: The plate is sprayed with 3.0 mL of the reagent using the Derivatizer (blue nozzle, spray level 3), then heated for 10 min at 100 °C on the TLC Plate Heater.
Documentation	Using the Visualizer 2, in UV 254 nm (Detection A) prior to derivatization, and in UV 366 nm (Detection B) and white light (Detection C) after derivatization. Peak profiles were generated from images acquired in Detections B and C.
Densitometry	Densitometry was not performed; classification used peak profiles from images only.

2. System Suitability Test (SST)

SST (Detection A)

The integrated System Suitability Test (SST) tool in *visionCATS* was used to verify chromatographic migration performance prior to machine-learning analysis.

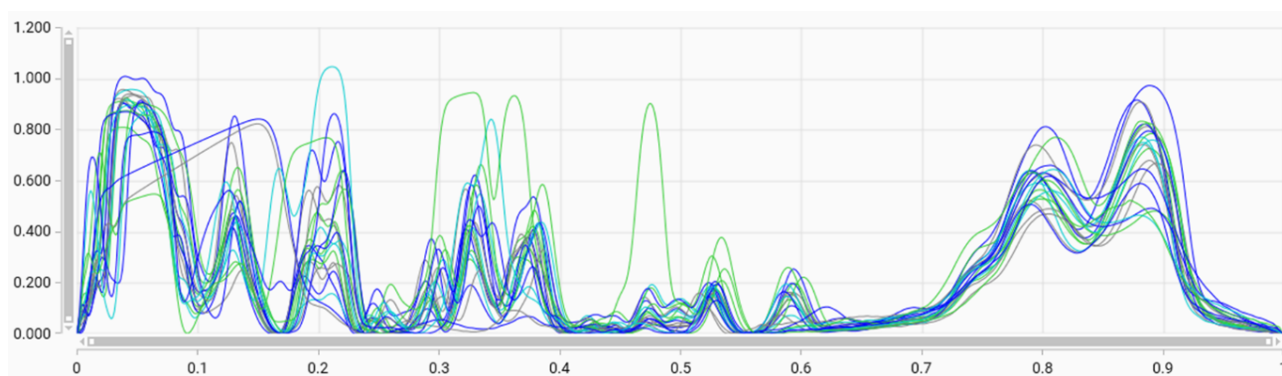
For each analysis, the UHM track was evaluated for the presence and positional conformity of five characteristic marker zones at $R_F \approx 0.193, 0.215, 0.494, 0.610$, and 0.765 . System suitability was confirmed when the deviation of each monitored zone was within $\Delta R_F \leq 0.01$.

3. Profile Generation

The Generate Profile tool in *visionCATS* was used to convert the acquired images into chromatographic profiles. For this application, profiles were generated from UV 366 nm and white-light images acquired after derivatization.

4. Comparison File Creation

The generated profiles were exported into a single comparison file, allowing all fingerprint profiles to be organized and analyzed together during the subsequent classification workflow.



Chromatographic profiles generated from images in UV 366 nm and white light after derivatization compiled in a single comparison file for subsequent classification analysis.

5. Classification using One-Class Support Vector Machine (OC-SVM)

The OC-SVM analysis was performed using the Classification tool integrated in *visionCATS* and accessed from the comparison file.

Unlike conventional classification approaches that require examples from multiple predefined classes, OC-SVM models only the distribution of the target reference class. During training, the algorithm learns a decision boundary in a multidimensional feature space that represents the authenticated reference samples. Unknown samples falling within this acceptance region are classified as conforming with the reference model, whereas samples falling outside the decision boundary are classified as non-conforming.

This approach is particularly suitable for authenticity assessment because authenticated reference materials are generally available, whereas representative examples of all possible adulterants, substitutes, degradation products, or other non-conforming materials are rarely available.

The following workflow was implemented using the guided classification workflow provided in *visionCATS*:

1. Classification Mode

In *visionCATS*, the supervised-learning workflow was selected to define the authenticated target class. Profiles from 23 authenticated red Asian ginseng samples were assigned to the target class and used for OC-SVM model training. The remaining 11 samples were assigned to the Unknown class: 10 had predefined expected status, while one sample had unresolved identity and subsequently evaluated against the trained model.

2. Data Preprocessing

The following preprocessing settings were applied:

- *Wavelength/Illumination Selection*: Profiles generated from UV 366 nm and white-light images after derivatization were selected.
- *Data Mode*: Profile mode was used so that the complete chromatographic fingerprint was included in the analysis rather than selected peaks only.
- *Profile processing*: Default *visionCATS* settings were applied for bounds selection, baseline correction, and normalization, without additional manual adjustment.
- *Binning*: A bin width of R_F 0.02 was applied to reduce the influence of minor positional variations between chromatographic profiles while retaining the overall fingerprint information.

3. Dimensionality Reduction

visionCATS provides several dimensionality-reduction techniques, including PCA, Kernel PCA, Sparse PCA, UMAP, and t-SNE. For this application, no dimensionality reduction was applied. The OC-SVM model was therefore trained using the full set of preprocessed and binned fingerprint variables.

4. Modeling

The OC-SVM algorithm was selected, and the Radial Basis Function (RBF) kernel was applied to enable the model to capture non-linear relationships within the multidimensional HPTLC fingerprint data. The default *visionCATS* settings for the following hyperparameters were used:

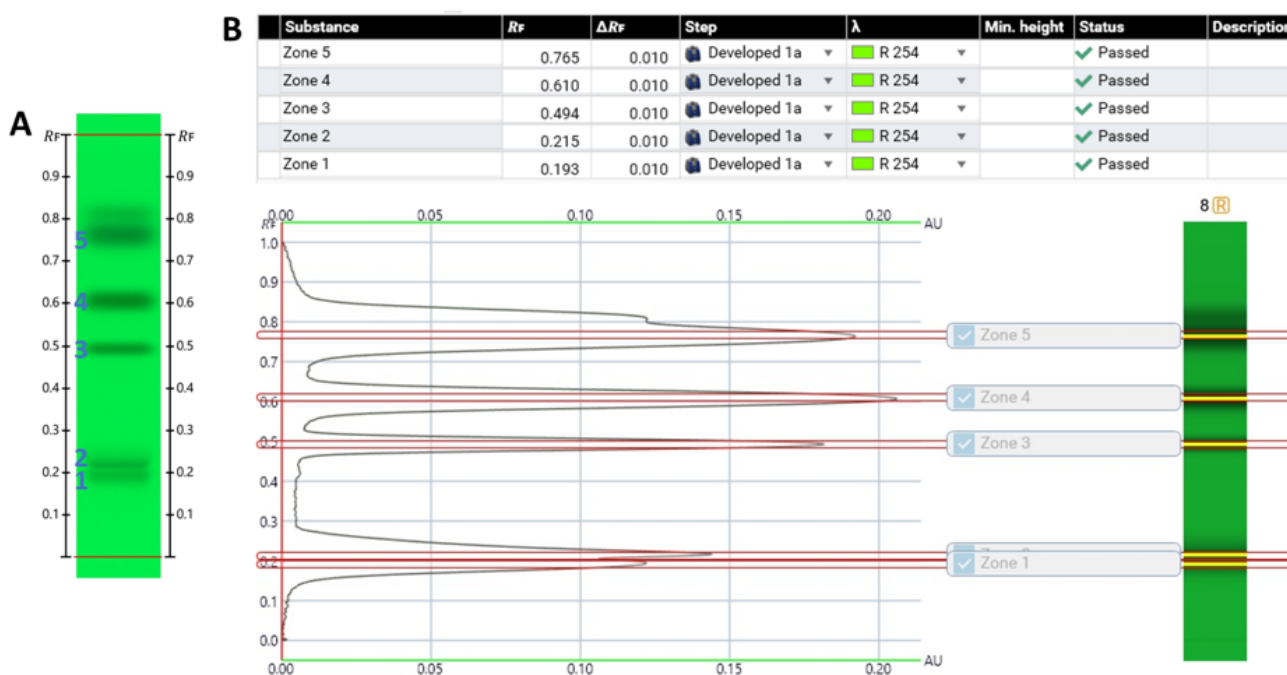
- Confidence Limit: 95%
- Nu: 0.1

The selected parameters were kept constant throughout the analysis and were not modified based on the classification results of the unknown samples.

6. Results

System Suitability Test

All chromatographic runs met the predefined SST acceptance criterion of $\Delta R_F \leq 0.01$ for the five monitored marker zones, confirming adequate reproducibility of chromatographic migration prior to machine-learning analysis.



System Suitability Test (SST) results obtained using the integrated visionCATS SST tool. The five monitored marker zones met the predefined positional acceptance criterion of $\Delta R_F \leq 0.01$, demonstrating consistent chromatographic migration across the analyzed runs.

Authenticity Assessment Using OC-SVM

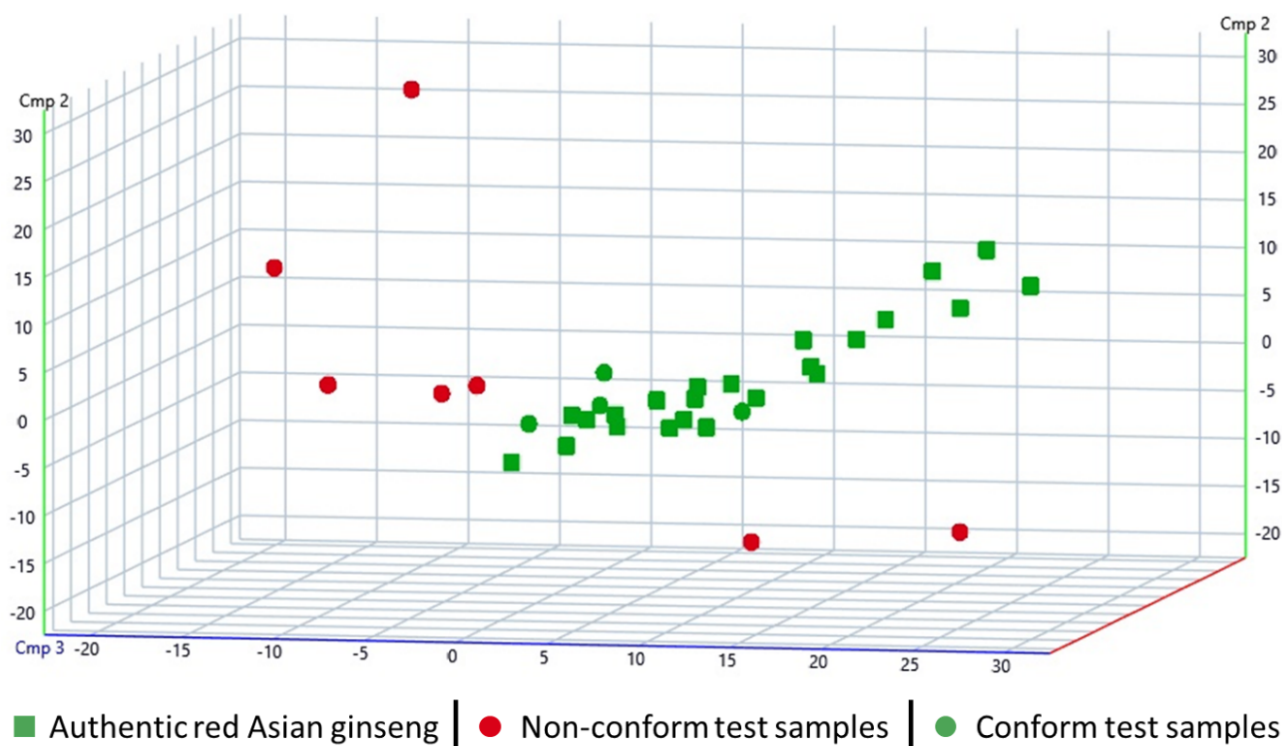
The model was trained exclusively using fingerprint data from the authenticated red Asian ginseng samples studied in this note, therefore, the model represents the variability captured by this specific reference set. During training, the OC-SVM established a multidimensional acceptance region representing the variability observed within the authenticated reference set.

For visualization of the classification results, visionCATS provides a three-dimensional PCA (3D PCA) plot. In this application note, the PCA plot is presented solely as a visualization tool to facilitate interpretation of the sample distribution and classification outcomes. The PCA scores were not used to train the OC-SVM model. Instead, the model was trained on the complete preprocessed and binned fingerprint dataset, without dimensionality reduction.

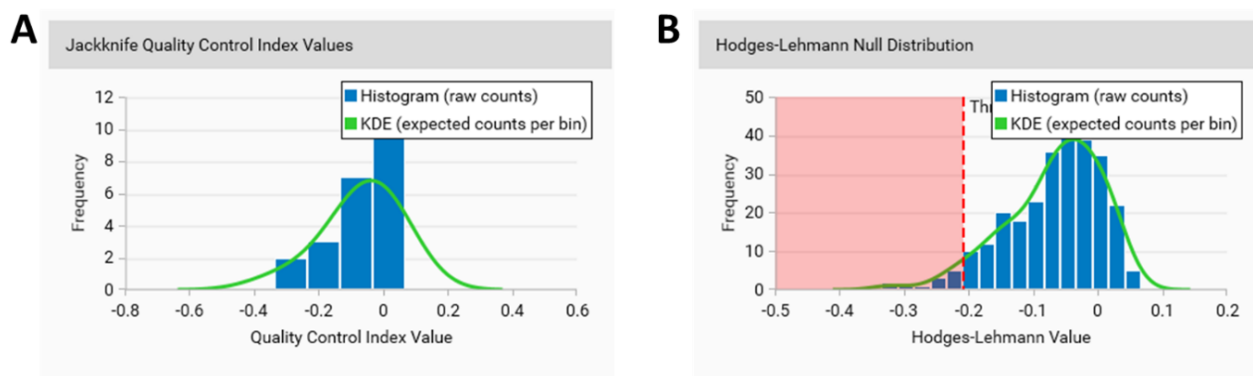
Unknown group was subsequently evaluated against this reference model. Rather than assigning samples to multiple predefined classes, the OC-SVM determines whether the fingerprint of an unknown sample falls within or outside the acceptance region defined by the authenticated reference samples.

To evaluate the internal consistency of the reference model, visionCATS provides a Jackknife Quality Control (QC) Index plot. This plot summarizes the results of Jackknife cross-validation and displays the distribution of QC index values obtained for the authenticated reference samples.

The Hodges-Lehmann Null Distribution plot represents the null distribution of the Hodges-Lehmann statistic and defines the threshold used for outlier detection within the model. Samples exceeding this threshold, represented by the red-shaded region, are considered inconsistent with the authenticated reference model and are classified as non-conforming.



Authenticity assessment of red Asian ginseng using OC-SVM classification in visionCATS. The figure displays the classification results in a 3D PCA plot. The PCA plot is only for visualization purposes to illustrate the distribution of conforming and non-conforming samples following classification and was not used for OC-SVM model training.



(A) The Jackknife Quality Control (QC) Index summarizes cross-validation results for the authenticated red Asian ginseng reference samples and demonstrates the internal consistency of the reference model. (B) The Hodges-Lehmann Null Distribution plot defines the statistical threshold used for outlier detection and classification of non-conforming samples.

The final classification results are summarized in a score table. Samples meeting the acceptance criteria defined by the OC-SVM model are highlighted in green, whereas samples exceeding the outlier threshold are highlighted in red. Of the 11 unknown samples evaluated, 4 were classified as conforming with the authenticated reference model and 7 as non-conforming. The classification scores provide a quantitative indication of each sample's position relative to the model acceptance threshold. Subsequent verification using known sample identities demonstrated agreement between the classification outcomes and the known botanical identity and processing type of the evaluated materials.

Results (Supervised)		
		Score
7	S26298-240828-01	-0.07
12	S26306-240828-01	-0.05
15	S26309-240828-01	-0.07
16	S26312-240828-01	-0.03
17	S26315-240828-01	-0.22
18	S26316-240828-01	-0.22
22	S22857-240902-01	-0.22
23	S22852-240902-01	-0.22
24	S18290-240902-01	-0.22
25	S14408-240807-01	-0.22
26	S22849-240902-01	-0.22

Classification results for eleven unknown samples evaluated using the trained OC-SVM model. Samples falling within the acceptance region defined by the authenticated red Asian ginseng reference set are highlighted in green and classified as conforming, whereas samples falling outside this acceptance region are highlighted in red and classified as non-conforming.

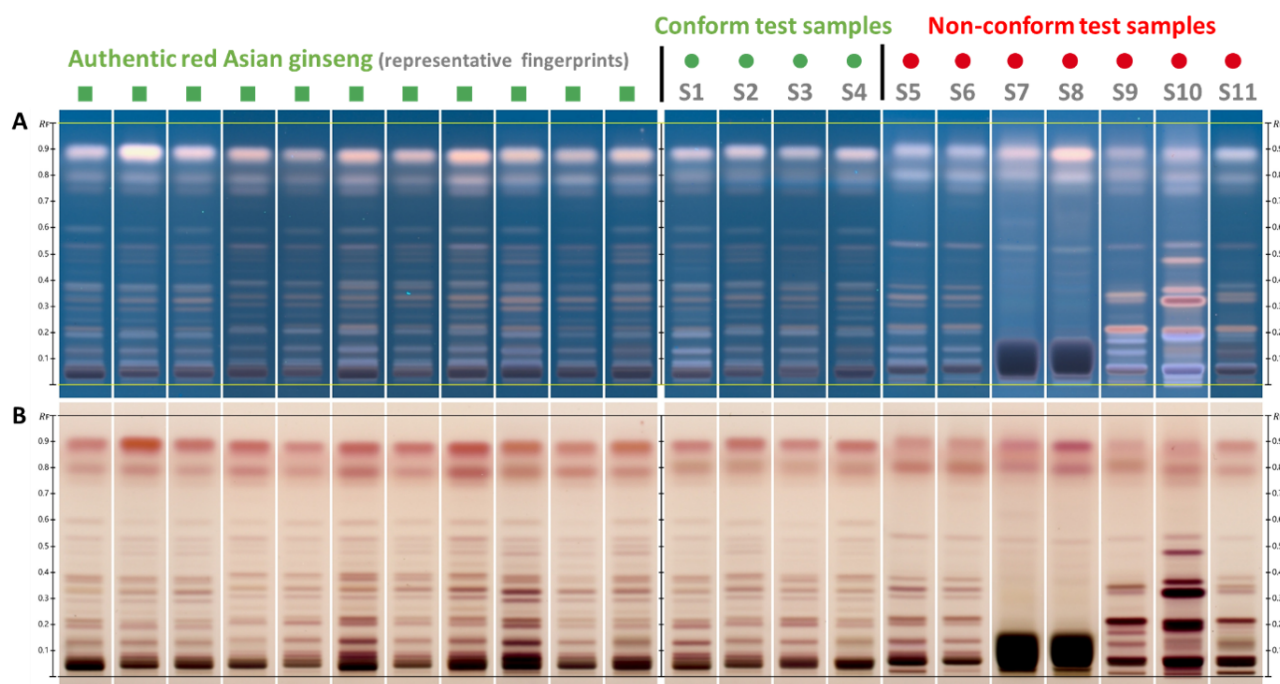
Verification of OC-SVM Classification Results

The known identity and expected status of the unknown samples were compared with the OC-SVM classification results to assess the agreement between the model predictions and the sample type. The OC-SVM classification results showed complete agreement with the known identity and expected status of the evaluated samples. Samples representing authentic red Asian ginseng were classified as conforming, whereas samples not belonging to the defined red Asian ginseng target class were classified as non-conforming.

Verification of the OC-SVM classification results. Comparison between the known identity of the unknown samples and the classification outcome obtained using the OC-SVM model.

Sample	Known Identity	Expected Status	OC-SVM Result
S1 S26298	Red Asian ginseng	Authentic	Conforming
S2 S26306	Red Asian ginseng	Authentic	Conforming
S3 S26309	Red Asian ginseng	Authentic	Conforming
S4 S26312	Red Asian ginseng	Authentic	Conforming
S5 S26315	White Asian ginseng	Non-conforming	Non-conforming
S6 S26316	White Asian ginseng	Non-conforming	Non-conforming
S7 S22857	<i>Codonopsis pilosula</i>	Non-conforming	Non-conforming
S8 S22852	<i>Codonopsis pilosula</i>	Non-conforming	Non-conforming
S9 S18290	American ginseng (<i>Panax quinquefolius</i>)	Non-conforming	Non-conforming
S10 S14408	<i>Panax notoginseng</i>	Non-conforming	Non-conforming
S11 S22849	Asian ginseng (type not specified)	Undetermined	Non-conforming

Visual inspection of the chromatographic fingerprints supported the OC-SVM classification results. Samples classified as non-conforming exhibited visually noticeable differences relative to the authenticated reference fingerprint pattern.



HPTLC fingerprints of the eleven unknown samples compared with authenticated red Asian ginseng reference samples. Images collected after derivatization in UV 366 nm (A) and white light (B). Several samples classified as non-conforming exhibited visually noticeable differences relative to the authenticated reference fingerprint pattern.

Conclusion

This application note demonstrates the application of the One-Class Support Vector Machine (OC-SVM) algorithm integrated in CAMAG *visionCATS* for the assessment of red Asian ginseng using HPTLC fingerprint data. During training, the OC-SVM established a multidimensional acceptance region representing the variability observed within the authenticated reference set.

The implemented workflow combines HPTLC fingerprinting, data preprocessing, and machine-learning-based one-class classification within a single software environment. In this study, the trained model differentiated unknown samples whose fingerprints were consistent with the authenticated reference set from samples falling outside the defined acceptance region. Comparison with the known identity and expected status of the evaluated samples showed agreement with the OC-SVM classification for all samples with a predefined expected status. Visual inspection of the HPTLC fingerprints further supported the classification results.

These results demonstrate the potential of the OC-SVM approach for identifying samples that do not conform to a defined authenticated reference class. Further evaluation using larger and more diverse authenticated reference sets and independent validation samples would allow the performance and robustness of the model to be more comprehensively established.