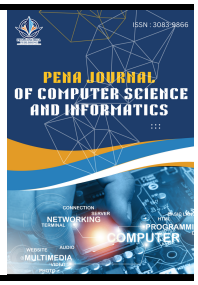
PENA CENDEKIA PUBLISHING
NURTURING IDEAS, ENRICHING KNOWLEDGE

Pena Journal of Computer Science and Informatics

Journal homepage:
https://penacendekia.com.my/index.php/pjcsi/index
ISSN: 3083-9866



Handheld REVA Spectrometer with Embedded Machine Learning Models for Rapid Stingless Bee Honey Authentication: A Field-Deployable Solution

Hanys Syazwana Harun[1,*], Abdur Rehman Laili[1], Zalhan Md Yusof[1], Nur Ain Mohd Aziz[1], Duratul Ain Rozli[1], Nur Irdina Hazizul[1]

[1] Photonics Technology Lab, MIMOS Berhad, MRANTI Park, 57000, Federal Territory of Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 28 July 2025
Received in revised form 14 August 2025
Accepted 10 September 2025
Available online 28 October 2025

Keywords:
REVA Spectrometer; portable Vis-NIR spectroscopy; embedded machine learning; honey adulteration; portable food authentication; One-Vs-Rest (OVR) Classification; Principal Component Analysis (PCA); Partial Least Square Regression (PLSR)

ABSTRACT

Honey is one of the most adulterated foods that caused the global issue in food fraud. Stingless Bee Honey (SBH), prized in ASEAN countries for its superior medical properties, commands premium prices and is prone to adulteration. Conventional authentication methods, though accurate, are time-consuming, costly and require specialized facilities, limiting their use in routine quality control.This study presents a rapid and non-destructive method by introducing a handheld Vis-NIR spectrometer (REVA In-Vitro) integrated with embedded machine learning models, operating in the 400-1000 nm range to authenticate SBH. Samples were artificially adulterated with distilled water (DW), apple cider vinegar (ACV), and fructose syrup (FS) at honey purity levels from 10% to 90% in 10% increments. 700 spectral data were used to train One-Vs-Rest (OVR) classification models with Principal Component Analysis (PCA) to classify adulterations, and Partial Least Square Regression (PLSR) for SBH purity detection. Validation was conducted using 50 unseen spectral data.The OVR classification model with PCA achieved 100% accuracy for both training and validation, while PLSR models attained determination coefficients, $R^2 > 0.999$ and root mean square errors, RMSE of 0.27% - 0.77% during training, and $R^2 \geq 0.938$ in validation. This approach offers rapid analysis, minimal sample preparation, and cost-effectiveness compared to conventional techniques, providing a practical solution for real-time SBH authentication in field settings.

1. Introduction

Stingless bee honey (SBH) represents a unique and economically valuable product in tropical and subtropical regions, particularly across Southeast Asian countries including Malaysia, Thailand, Indonesia, and the Philippines. SBH is produced by stingless bees (Meliponini) and is characterized by distinct physicochemical properties, including higher moisture content (typically >20%) and enhanced therapeutic properties, which is different from conventional honey from *Apis mellifera* [1].

* *Corresponding author.*
E-mail address: *hanys.harun@mimos.my*

https://doi.org/10.37934/pjcsi.2.1.4766

The high market value of SBH creates both opportunities and challenges for the industry. With prices 3-10 times higher than regular honey, premium SBH reflects consumer recognition of its superior medicinal properties and limited production volumes. However, the high prices also created significant economic incentives for adulteration, threatening both consumer trust and the sustainability of traditional meliponiculture practitioners.

The authentication of honey, particularly premium varieties like SBH, has consequently become a critical challenge for the global food industry. Economic adulteration through the addition of cheaper sweeteners such as high fructose corn syrup, rice syrup, and other sugar solutions represents a widespread problem that undermines market integrity [2]. This issue is particularly acute for SBH, where the high market value makes fraudulent practices economically attractive to unscrupulous actors.

Current analytical methods present significant barriers for routine quality control implementation even though they provide accurate results. Techniques such as high-performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR) spectroscopy, and isotope ratio mass spectrometry (IRMS) require advanced laboratory infrastructure, specialized expertise, and significant time investment. For example, while NMR can provide long-term stability of spectra and extensive structural information, it requires costly instrumentation and remains inaccessible to most honey producers [3]. Similarly, even though HPLC is effective for sugar composition analysis, it involves complex sample preparation and can fail to detect sophisticated adulterants that resemble natural honey composition.

These limitations have created a significant research and practical gap: the honey industry urgently needs rapid, accessible, and cost-effective authentication methods that can be deployed outside traditional laboratory settings. This gap is particularly pronounced for SBH producers and traders who often operate in remote locations with limited access to analytical facilities.

Addressing this gap holds significant scientific and practical importance for multiple stakeholders. From a scientific perspective, developing portable authentication methods for SBH requires novel approaches that account for its unique compositional characteristics, which differ substantially from conventional honey. This research advances portable analytical chemistry field and demonstrates the integration of machine learning with field-deployable instrumentation. From a practical standpoint, this research addresses critical industry needs by potentially democratizing quality control, protecting premium market integrity, and supporting sustainable livelihoods for traditional meliponiculture practitioners. The significance extends beyond individual producers, contributing to consumer protection, regulatory enforcement, and the preservation of authentic food products in global markets.

Near-infrared (NIR) spectroscopy, especially when integrated with machine learning algorithms and embedded in portable platforms, represents a promising solution to the identified gap. This technique offers significant advantages such as rapid analysis, cost-effectiveness, minimal sample preparation, and non-destructive analysis capabilities [4]. Recent developments in portable NIR technology have shown potential for on-site analysis without specialized laboratory facilities [5]. However, most existing studies on honey authentication using spectroscopy have centered on conventional honey types, with limited focus on the unique compositional characteristics of SBH that require specialized analytical approaches [8].

This study presents the development and validation of a rapid, non-destructive method for detecting and quantifying adulteration in stingless bee honey using the REVA In-Vitro spectrometer, a portable Vis-NIR platform operating in the 400-1000 nm range. This approach directly addresses the identified research gap by combining One-Vs-Rest (OVR) classification employed with principal

component analysis (PCA) for adulterant classification with partial least squares regression (PLSR) for quantitative honey purity determination.

The specific objectives of this research are to: (1) develop robust classification models capable of identifying common adulterants in SBH including distilled water, apple cider vinegar, and fructose syrup; (2) establish quantitative prediction models for determining honey purity levels with high accuracy; (3) validate the performance of these embedded models using independent samples; and (4) demonstrate the practical applicability of portable Vis-NIR spectroscopy for routine quality control in SBH production and trade.

This work aims to deliver an accessible and cost-effective solution that empowers stakeholders across the honey supply chain, from small-scale producers to regulatory authorities. By bridging the gap between laboratory-grade analysis and field-deployable technology, this research ultimately seeks to protect consumer interests while supporting the sustainable growth of the SBH industry.

2. Methodology

2.1 Instrumentation and Spectral Acquisition

This section outlines the instrumentation and protocols used for spectral data acquisition. The primary analytical instrument was a REVA In-Vitro Vis-NIR spectrometer for honey authentication measurements, supplemented by a digital refractometer for independent verification of sample composition through Brix value determination.

2.1.1 Primary spectroscopic equipment

Spectral measurements were conducted using a Reagentless Blood Testing and Vital Sign Analyzer In-Vitro (REVA In-Vitro), a miniature Vis-NIR spectrometer operating in the 400-1000 nm range that covers visible to near-infrared spectroscopy as shown Figure 1 which illustrates its compactness compared to a standard computer mouse. A palm size REVA In-Vitro, with dimensions of 119 mm (length) × 74 mm (height) × 42 mm (width) is a highly portable device and suitable for field applications or any on-site spectral analysis. This portable platform serves as the primary analytical instrument for honey authentication, offering the capability to embed machine learning models directly into the device for real-time field analysis.

Samples were placed in 10 × 10 mm glass cuvettes and scanned at room temperature. For each mixture, spectral acquisition followed a systematic protocol where five individual scans were performed, with each scan generating five spectra, resulting in 25 spectra per sample. This approach ensured adequate sampling for statistical reliability while capturing potential measurement variability.

Fig. 1. REVA In-Vitro, a mini spectrometer compared to computer mouse

2.1.2 Supplementary equipment

A digital refractometer HI 96803 from Hanna Instrument was employed as supplementary instrumentation to measure Brix values from all samples, providing independent verification of sugar content and compositional changes during sample preparation as shown in Figure 2.

Fig. 2. Digital refractometer HI 96803

2.2 Sample Preparation

Stingless bee honey (SBH) samples were procured from local producers. To systematically evaluate adulteration detection capabilities, adulterants such as distilled water (DW), apple cider vinegar (ACV), and fructose syrup (FS) were mixed with pure honey to create samples with purity levels ranging from 10% to 90% in 10% increments, as shown in Table 1. Additional validation samples at 20%, 50%, and 80% honey purity levels for all classes of adulterated samples were prepared from a new batch to validate the trained models.

Table 1
Class of adulterated samples

Class	Adulterated samples
DW	SBH + distilled water
ACV	SBH + apple cider vinegar
FS	SBH + fructose syrup
H	SBH

Brix values (°Bx) were measured using the digital refractometer for each honey purity level ranging from 10% to 90% for all three classes of adulterants, as well as for pure honey (100%), as shown in Table 2. Brix measurements served as an independent quality verification method to confirm successful sample preparation and to provide supplementary compositional data for future correlation studies. While these values were not incorporated into the machine learning model development due to the focus on developing a purely spectroscopic approach, they provided valuable confirmation of the expected compositional changes in the adulterated samples and served as a quality control measure during the experimental phase.

Table 2
Brix values (°Bx) for all samples in different honey purity level (%) using a digital refractometer

Honey purity level (%)	10	20	30	40	50	60	70	80	90	100 (Pure Honey)
					Brix (°Bx)					
DW	8.9	20	24.3	32	40.1	45.2	50.8	56.4	62.6	70.9
ACV	14.8	19.8	28.5	32.9	40.2	43.4	49.9	58.7	64.9	
FS	76.5	76.1	75.7	73.7	74.4	73.7	72.7	72.6	72.6	

The primary analytical instrument employed was a Reagentless Blood Testing and Vital Sign Analyzer In-Vitro (REVA In-Vitro), a miniature and portable Vis-NIR spectrometer since the machine learning models can be embedded directly into the device, enabling real-time analysis without external computational resources. This device represents the core innovation of this study, providing both spectral acquisition and on-device analytical processing capabilities.

A digital refractometer was employed as a supplementary analytical tool to measure Brix values from all prepared samples. This instrument served dual purposes which is providing independent verification of successful sample preparation by confirming expected compositional changes and generating complementary data for potential future correlation studies with spectroscopic measurements.

2.3 Data Collection and Preprocessing

This section details the systematic approach used for dataset generation and spectral data preparation for the model development from the machine learning. The methodology encompasses the collection of a comprehensive training dataset comprising 700 spectra from various honey-adulterant mixtures, alongside an independent validation dataset of 50 spectra for model performance assessment.

2.3.1 Dataset composition

A total of 700 spectra were collected for model training. This dataset was generated from 28 distinct honey-adulterant mixtures (representing purity levels from 10% to 90% in 10% increments across the three adulterant classes, plus pure honey controls). Each mixture underwent 5 individual scans, with each scan producing 5 spectral measurements, yielding the complete training dataset (28 mixtures × 5 scans × 5 spectra = 700 total spectra).

For model validation, 50 unseen spectra were acquired from newly prepared samples at three specific purity levels (20%, 50%, and 80%) across all adulterant classes and this provides assessment of model performance on new data representative of practical application scenarios.

2.3.2 Spectral preprocessing

Standardization (Standard Scaler) preprocessing was applied to raw absorbance values from the REVA In-Vitro to remove spectral noise and normalize data for both classification and regression model development. Standard Scaler implements Z-score normalization, a well-established and commonly used scaling technique that centers data around zero with unit variance [12].

2.4 One-Vs-Rest (OVR) Classification Model with PCA Employed

The One-Vs-Rest (OVR) Classifier was implemented for adulterant type identification, selected as the optimal algorithm for classification accuracy [13]. Principal Component Analysis (PCA), which is one of the dimension reduction techniques, was employed to reduce the dimensionality of spectral data and facilitate efficient classification of adulterant types.

Model training incorporated parameter optimization through Grid-Search with 5-fold cross-validation (K-Fold Stratified) to optimize OVR parameters and ensure robust performance. The classification objective was to evaluate sample clustering according to adulterant mixture types, with model performance assessed through accuracy metrics, confusion matrices, and 2D/3D score plots during training using the 700 spectra dataset. Model validation employed 50 unseen spectra to evaluate accuracy and confusion matrices for association strength verification.

2.5 Partial Least Square Regression (PLSR) Model

Separate PLSR models were developed for each adulterant type (DW, ACV, FS) to predict honey purity levels quantitatively. Model training utilized parameter optimization through Grid-Search with 5-fold cross-validation (K-Fold Stratified) to optimize PLSR parameters and ensure model robustness.

Model performance evaluation involves determination coefficient (R^2), root mean square error (RMSE), and linear regression analysis of actual versus predicted honey purity levels for both training and validation datasets. Regression coefficient plots were generated during model training to identify significant wavelengths contributing to each model, providing insight into the spectral regions most relevant for purity determination in each adulterant class.

3. Results

3.1 Absorbance Data

The Vis-NIR absorbance spectra of SBH adulterated with distilled water (DW), apple cider vinegar (ACV), and fructose syrup (FS) show distinct profiles across the wavelength range of 415–940 nm, reflecting the impact of different adulterants on the optical properties of honey.

As illustrated in Figure 3, samples with higher concentrations of DW exhibit markedly higher absorbance intensities compared to the pure SBH of 100% with the absence of DW and those with lower adulteration levels. This observation of inverse relationships is counterintuitive, as dilution is generally expected to reduce absorbance, but it may be resulted from multiple scattering, baseline shift, or interactions involving O–H stretching overtones and combination bands of water that overlap with broad organic absorbance features. Shifting vibrational modes and altering absorbance intensity happen when the water molecules form the hydrogen bonds with the constituents in honey. Previous study reported that the determination of water content in products is quite challenging since the absorption of water in Vis-NIR region from 300 nm to 1000 nm is very weak but Vis-NIR spectroscopy to determine water content is feasible since water is a primary quality parameter [9].

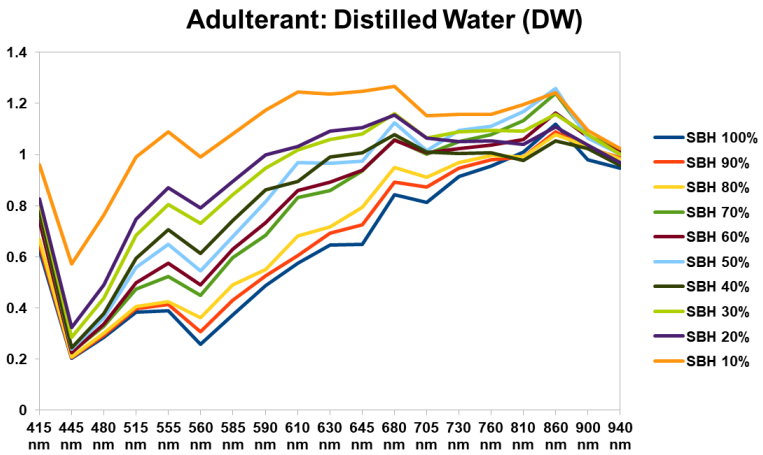


Fig. 3. Absorbance spectra for SBH samples that adulterated with distilled water (DW)

In contrast, the absorbance curves for ACV-adulterated samples in Figure 4 show a clear upward trend in absorbance that is visible from 415 nm to around 860 nm across all concentrations, with less pronounced separation between levels, indicating minimal variability induced by ACV in the Vis-NIR region. This behavior suggests it may be caused by the chemical effect which likely arises from acetic acid and O–H and aromatic bonds present in phenolic compounds in ACV that are known to contribute to absorbance in the Vis-NIR range. Previous study using hyperspectral imaging clarified that wavelength 400 nm to 700 nm showing sensitivity to certain phenolic compounds in ACV which aligned with a clear upward trend in the absorbance graph as shown in Figure 4 [11].

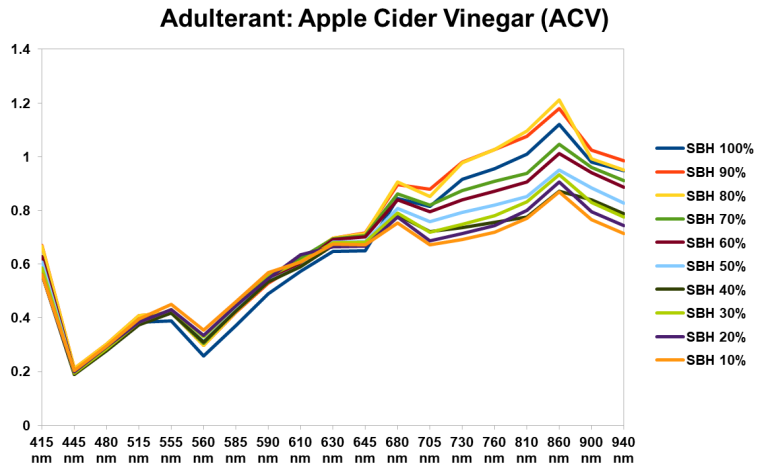


Fig. 4. Absorbance spectra for SBH samples that adulterated with apple cider vinegar (ACV)

FS-adulterated samples in Figure 5 show the most significant absorbance where the spectral behavior is highly discriminative, with clear separation between concentrations from 415 nm to 940 nm. This region, although outside the conventional NIR range, is known to capture second overtones of fundamental molecular vibrations. The observed absorbance progression is attributed to higher-order transitions of O–H and C–H bonds, characteristic of fructose, which are also responsible for strong absorption features in the NIR range. Figure 5 shows that all concentrations of FS exhibit the

highest peak at 860 nm, which closely aligns with a previous study reporting that one of the highest specificities for FS adulteration detection was found at 847 nm [8].

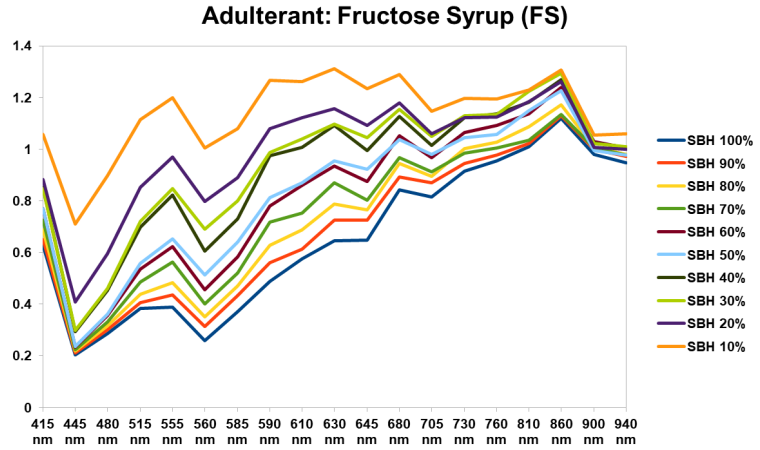


Fig. 5. Absorbance spectra for SBH samples that adulterated with fructose syrup (FS)

It is worth noticing that Brix values recorded during sample preparation as in Section 2.1 previously generally decreases with increasing adulteration levels for DW and ACV, while generally increases for FS, consistent with the expected dilution of the sugar content in the samples. Overall, the addition of water or apple cider, which is predominantly water, reduced the soluble solids content (SSC), whereas the addition of high fructose syrup led to a further increase in SSC [8]. Theoretically, this supports the spectral observation, particularly for samples that adulterated with FS where increasing sugar concentrations that reflected in higher Brix values align with the stronger absorbance features attributed to O–H and C–H bonds overtones. Although Brix values were not used in model training, their trends confirmed that there is a compositional change in the mixtures which shows that the spectral patterns are directly associated with the chemical content variations such as sugar. Overall, the spectral response varies significantly by adulterant type. These findings are discussed in Sections 3.2 and 3.3 below.

3.2 OVR Classification Model with PCA Employed

This section presents the performance of One-Vs-Rest (OVR) classification model that employed with Principal Component Analysis (PCA) for dimensionality reduction of the spectra to identify the type of adulterant present in SBH samples. The score of accuracy, precision, recall, and F1-score resulted to the model performance during both training and validation phases.

3.2.1 Training

The OVR classification model that employed with PCA was trained with the aim of classifying the honey samples into 4 classes and all those classes were successfully distinguished with 100% in overall accuracy, F1-score, precision and recall with the support of 225 for class DW, ACV and FS, and 25 for class H from the model trained on cuvette-based spectra data as shown in Table 3 and Figure 6. G.Mahalakshmi *et al*. (2021) reported that they achieved identical 100% accuracy for Random Forest among the seven classification algorithms for medical dataset which shows that it is possible to achieve 100% accuracy of classification [14].

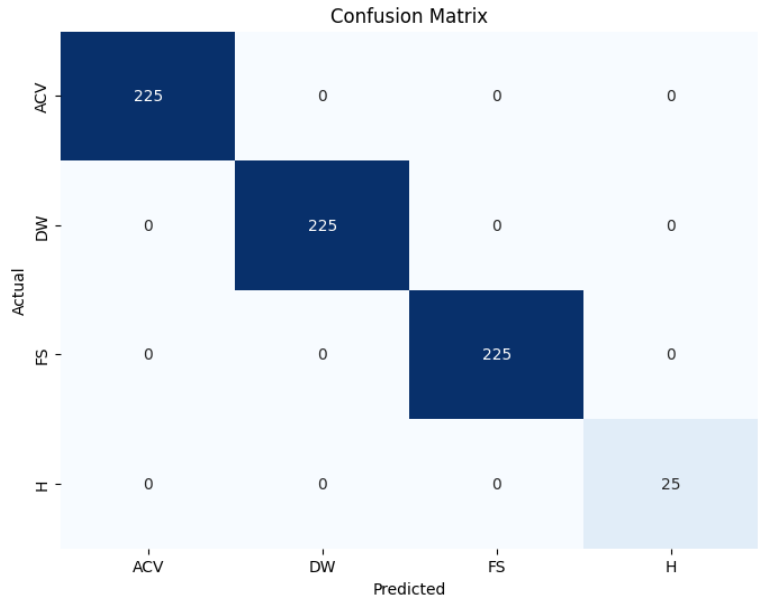


Fig. 6. Confusion matrix from the OVR classification with PCA for all 700 training samples

Table 3

Training result from the OVR classification with PCA model trained

Class of adulterated samples	Accuracy overall	F1-score	Precision	Recall	Support
DW	100	100	100	100	225
ACV		100	100	100	225
FS		100	100	100	225
H		100	100	100	25

During the training, cross validation using 5-fold (Stratified K-Fold) was applied via hyperparameter optimization of Grid Search to ensure the model's robustness where only the best model is saved and deployed. Proposed by [7] that Grid Search is carried out to optimize the parameters to enhance the efficiency of the model and concluded that hyper-parameter tuning plays a significant role and give a positive impact on the model's predictive power. The technique applied in this training is not only optimized for the model hyperparameter such as the principal component numbers but also to ensure that each class was fairly represented in all folds to reduce the risk of overfitting.

In our study, to simplify complex data like spectra into meaningful patterns, Principal Component Analysis (PCA) can be thought of to summarize the most important trends in the data which is similar to capturing the essence of a story without repeating every detail. PCA reduces data dimensionality by projecting them onto a lower-dimensional space which is principal components (PC) to summarize the data using only fewer components, and from this study [10], only first principal component (PC1) is retained and the rest are discarded. [6] observed that the first three PCs were used to reconstruct the spectra and confirmed that when additional PCs are added to PC1, the reconstructed spectra is better and matched the recorded data.

As shown in Figure 7, the 2D PCA score plot involved the principal components of PC1 and PC2 for the honey samples, SBH with total explained variance are 96.1 % and it visualizes moderate to good class separation, with some overlaps between classes. The 3D PCA score plot for the honey samples includes PC1, PC2 and PC3 with a total explained variance of 98.2% as shown in Figure 8. The inclusion of the third principal component provides an additional dimension of data separation, enhancing the distinction between classes that appear closer to the 2D plot. Those three components (PC1-PC3) are most likely to explain the vibration in the bonds of any adulterants which may contribute to their optical or chemical properties in the wavelength range of our study (415 nm - 940 nm). The PCA score plots indicate that PCA has effectively captured the main variance in the dataset such as optical or chemical properties in the honey, providing a simplified yet informative representation of the data which is meaningful for OVR classification.

The remaining 3.9% (2D) and 1.8% (3D) unexplained variance likely reflects subtle sample-specific variability, experimental noise, or minor compositional differences not crucial for class discrimination. These components may include minor baseline shifts, instrument noise, or low-variance chemical signatures which might be affected by adulterants such as water. Although the loading plots are not shown, the interpretation of PC1–PC3 follows prior spectral studies and is based on known Vis-NIR molecular vibration patterns.

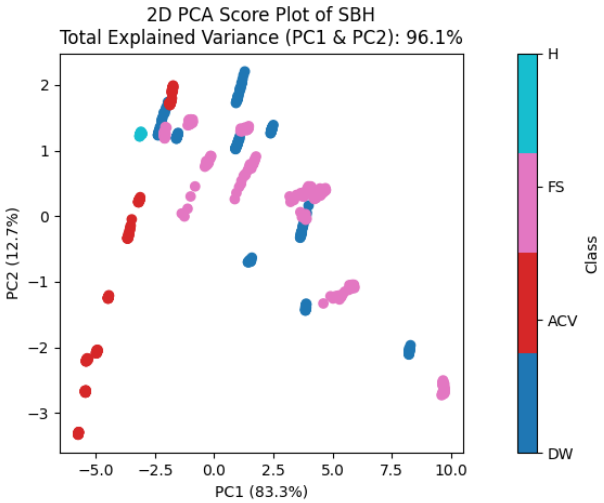


Fig. 7. 2D PCA score plots for all 700 training samples

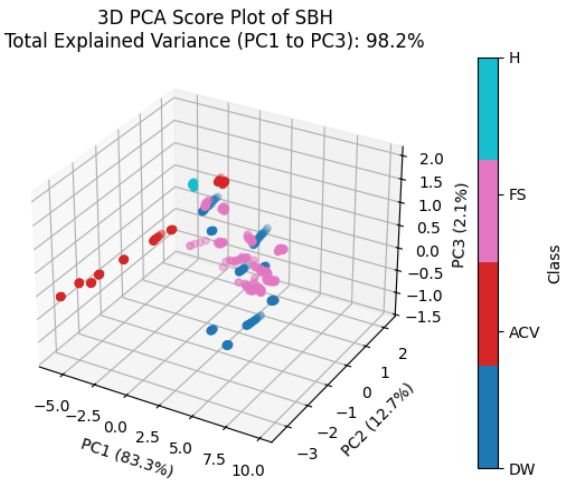


Fig. 8. 3D PCA score plots for all 700 training samples

Overall, the OVR classification model with PCA provides a clear and interpretable low-dimensional projection of the honey dataset. It captures meaningful chemical variance and supports reliable classification of SBH adulteration types.

3.2.2 Validation

To validate the OVR classification model with PCA employed, the model was tested on 50 new SBH samples with 15 samples for each of the adulterants DW, ACV and FS, while 5 samples for pure honey (H). As shown in Figure 9 and Table 4, all samples were correctly classified and again achieving 100% classification in overall accuracy, F1-score, precision and recall indicating a perfect association between actual and predicted classes even using the new data which is considered as unseen data in the real-world application, while this confirms the generalization capability of the OVR classification model with PCA when applied to unseen data.

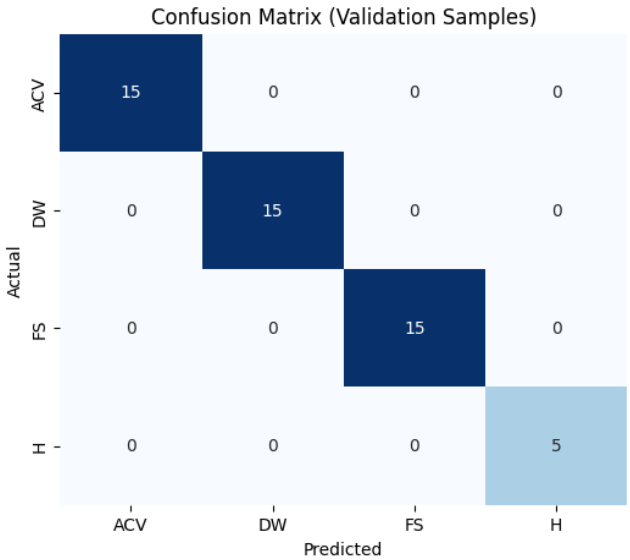


Fig. 9. Confusion matrix from the OVR classification model with PCA for all 50 validation samples

Table 4
Validation results from the OVR classification with PCA model trained

Class of adulterated samples	Accuracy overall	F1-score	Precision	Recall	Support
DW	100	100	100	100	15
ACV		100	100	100	15
FS		100	100	100	15
H		100	100	100	5

3.3 PLSR Regression Model

This section describes the performance of the Partial Least Squares Regression (PLSR) models developed to quantify the percentage of honey purity across three adulterant classes (DW, ACV, FS). The score R^2 and RMSE metrics contribute to the model performance for both training and unseen validation data, also regression coefficient analysis was performed to interpret the spectral significance of each wavelength during the training.

3.3.1 Training

Three PLSR models were developed for each of SBH adulterant classes, DW, ACV, and FS, to predict honey purity level (%) and the result display is combined with the classification result, such as 'DW, H: 30%' which indicates that the sample is a mixture of 70% distilled water and 30% honey. Those 3 models validate the honey level almost correctly for all new samples in the glass cuvette which have the same condition as the training data which will be discussed in the next part. Table 5 shows that all 3 PLSR models demonstrated excellent fit during training, with R^2 values above 0.999 and RMSE values ranging from 0.27 to 0.77, indicating high prediction accuracy across a broad range of purity levels (10% to 100%).

Table 5
Training result from the PLSR regression model trained

Class of Adulterated Samples	R^2	RMSE
DW	0.9999	0.27
ACV	0.9998	0.36
FS	0.9991	0.77

Same as the OVR classification model with PCA, the PLSR model also trained with the implementation of cross validation of 5-fold (Stratified K-Fold) via Grid Search to ensure the model robustness and promising accuracy. Regression coefficients plots identify the significant wavelengths contributing to each of the model, high positive coefficient indicates a strong positive influence that is crucial and critical for the predictions which may provide more insight into the chemical or physical properties of SBH represented in the models and vice versa for high negative coefficient. Wavelength with smaller or near-zero coefficients shows it has little influence on the model's prediction.

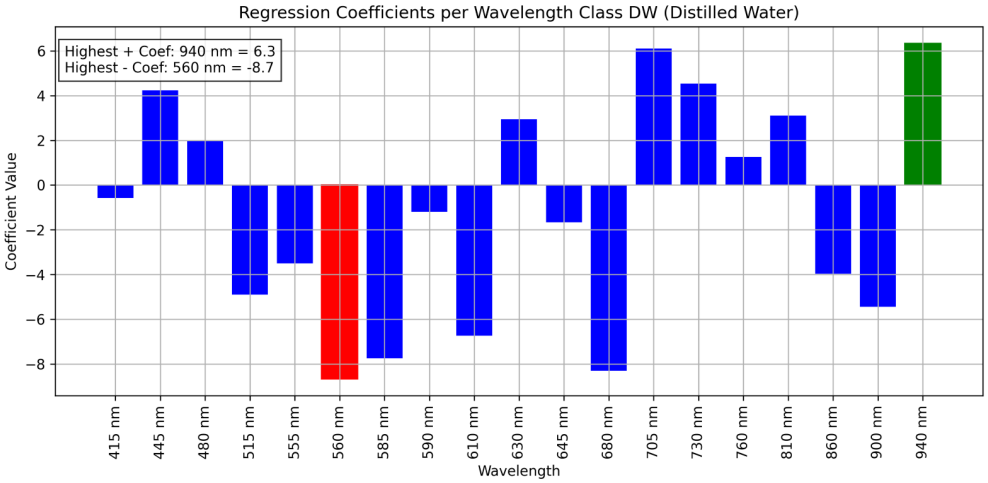
Table 6 and Figure 10 capture the result and plot of PLSR regression coefficient per wavelength for 3 classes of all samples. For PLSR model of DW-adulterated samples, the most influential wavelength is 940 nm, which corresponds to the O–H stretching overtone absorption of water. This region is known for strong NIR absorbance due to hydrogen bonding interactions of water molecules

with solutes, as described in aquaphotomic. In the ACV model, 705 nm emerged as the most positively influential, linked to overtone and combination bands associated with phenolic compounds in the ACV. For the FS model, the 760 nm wavelength shows the highest positive influence even though it is outside conventional NIR range, it still captures second overtone transitions of O–H and C–H bonds, primarily due to glucose and fructose content.

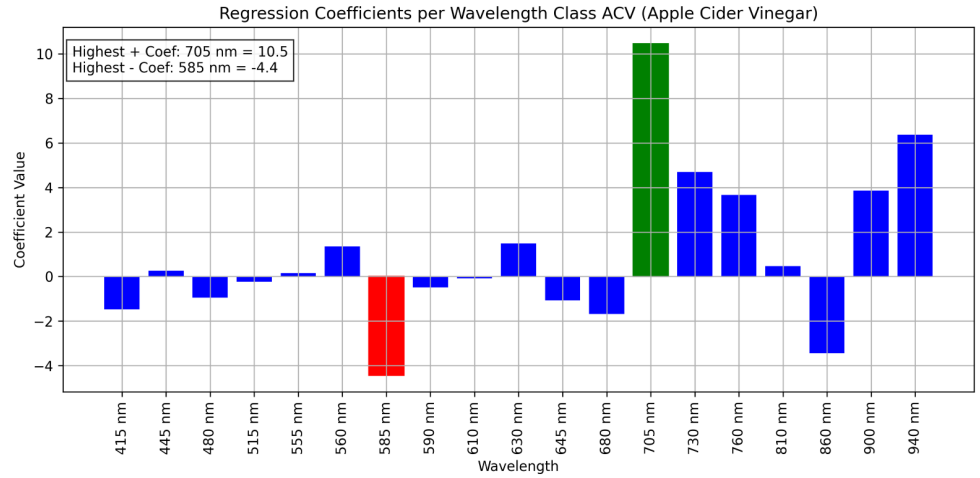
Table 6

PLSR regression coefficient per wavelength for 3 classes of all samples

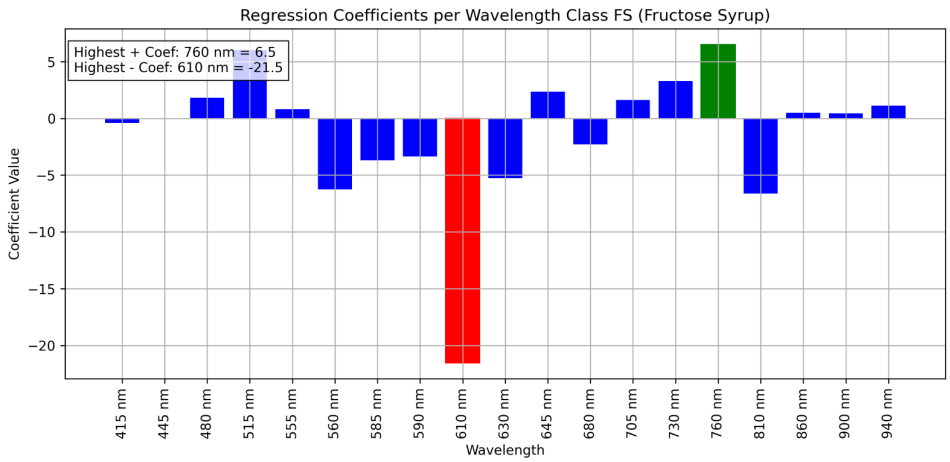
Class of Adulterated Samples	Highest Positive Coefficient		Highest Negative Coefficient	
	Wavelength (nm)	Coefficient	Wavelength (nm)	Coefficient
DW	940	6.3	560	-8.7
ACV	705	10.5	585	-4.4
FS	760	6.5	610	-21.5



(a)



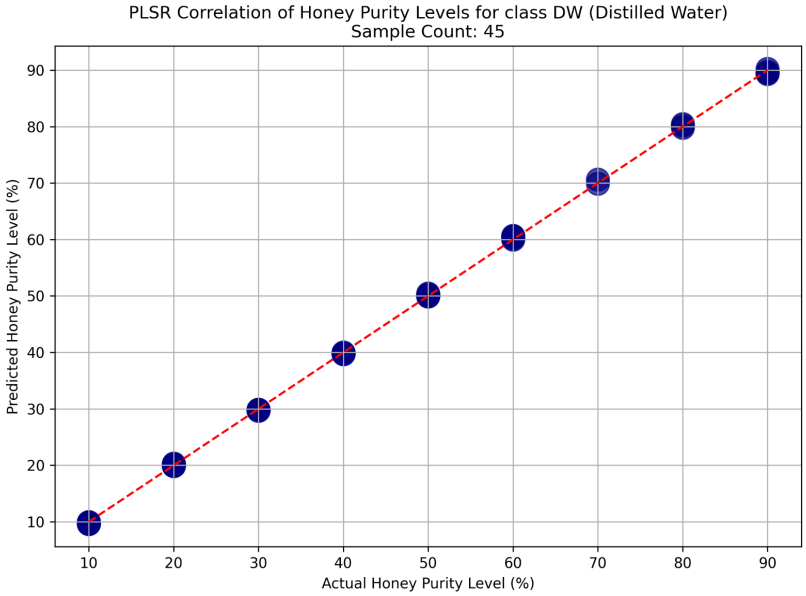
(b)



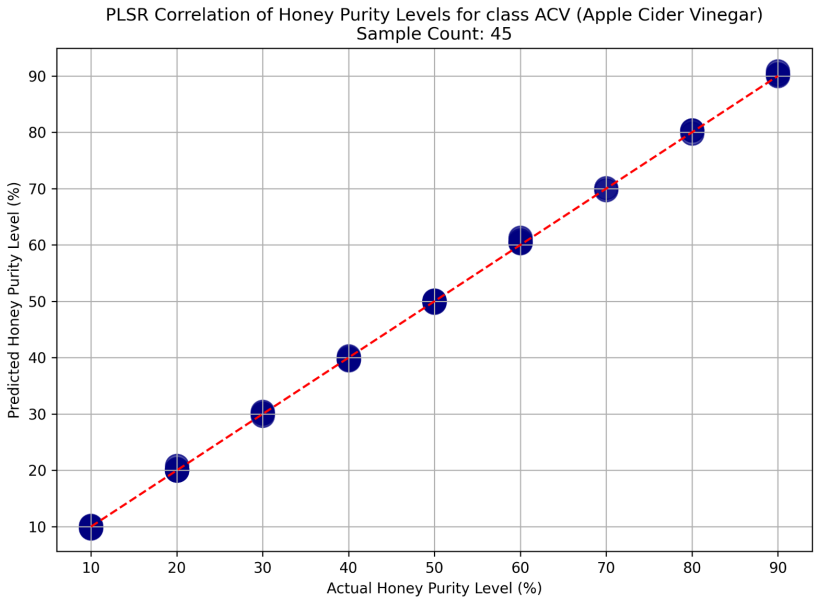
(c)

Fig. 10. Plot of PLSR regression coefficient per wavelength for samples adulterated with (a) distilled water (DW), (b) apple cider vinegar (ACV), and (c) fructose syrup

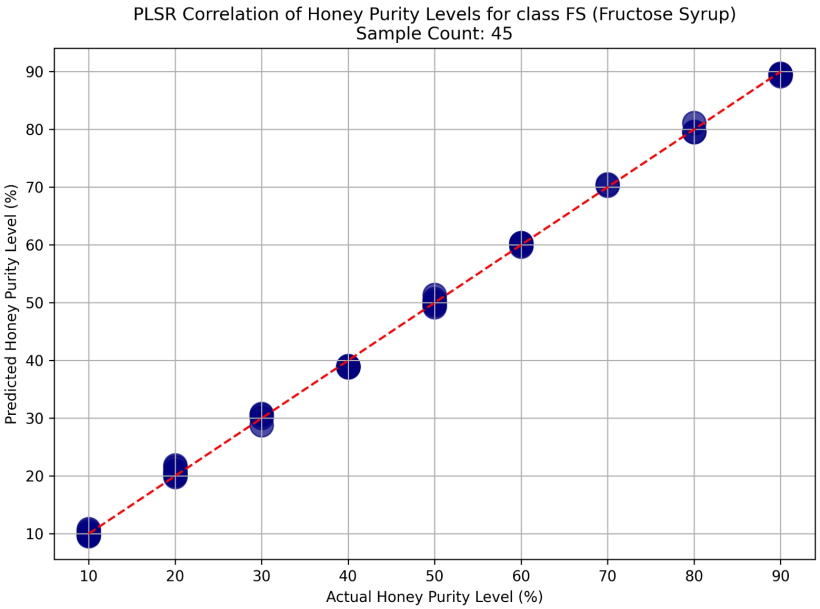
Linear regression of actual versus predicted honey purity level (%) of the 3 PLSR model determines the model's predictive capacity across the different classes (DW, ACV and FS) as in Figure 11. The predicted honey purity level (%) perfectly aligned along the diagonal (dashed red line) shows a high degree of linearity for each plot of all models. This alignment indicates the high accuracy of prediction, and the data points that are tightly clustered around the regression line proved the model's consistency.



(a)



(b)



(c)

Fig. 11. Plot of PLSR correlation of honey purity level between actual versus predicted sample that adulterated with (a) distilled water (DW), (b) apple cider vinegar (ACV), and (c) fructose syrup (FS)

3.3.2 Validation

For PLSR model validation, new SBH samples at three purity levels (20%, 50%, and 80%) were used to evaluate all trained models. This validation scheme, while reflective of practical purity extremes and a mid-point, introduces limitations in granularity and generalization, particularly since other purity levels which in SBH range of 10-100% were not included.

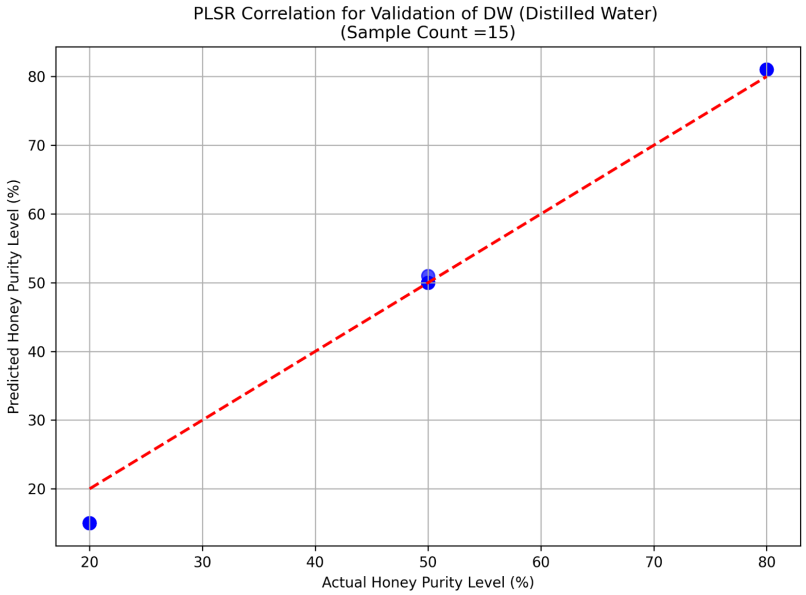
As expected, RMSE values increased compared to training, particularly for the FS model (RMSE = 6.10), due to reduced variation of concentration in data and the model's sensitivity to overlapping sugar-based spectral features as shown in Table 7. A similar observation was reported in a previous

study [15], where the model performance dropped from the reduced R^2 and increased normalize RMSE when tested on a limited testing data that did not reflect the full variability of seen data during training. Despite the increase of RMSE in our study, R^2 values remained high, ranging from 0.938 (FS) to 0.986 (ACV), showing that the models continue to capture the relationship between the absorbance and honey purity effectively.

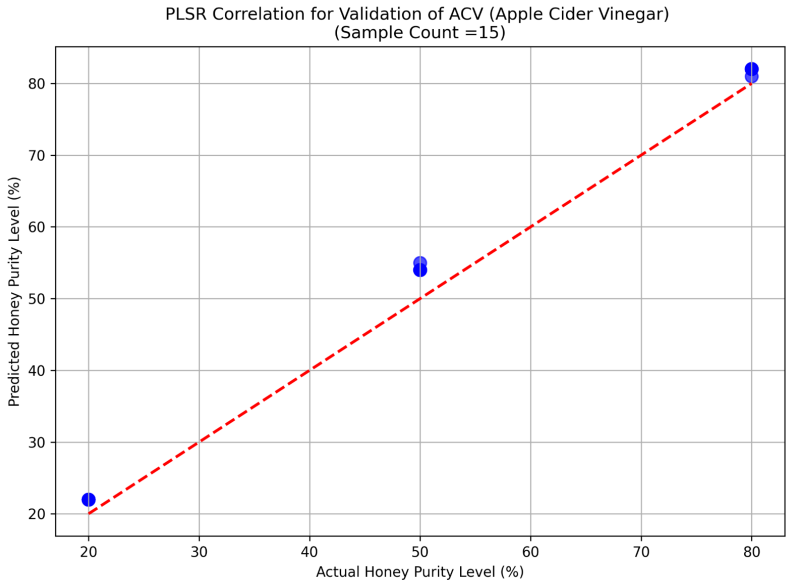
Table 7
Validation results from the PLSR regression model trained

Class of Adulterated Samples	R^2	RMSE
DW	0.9854	2.96
ACV	0.9860	2.90
FS	0.9380	6.10

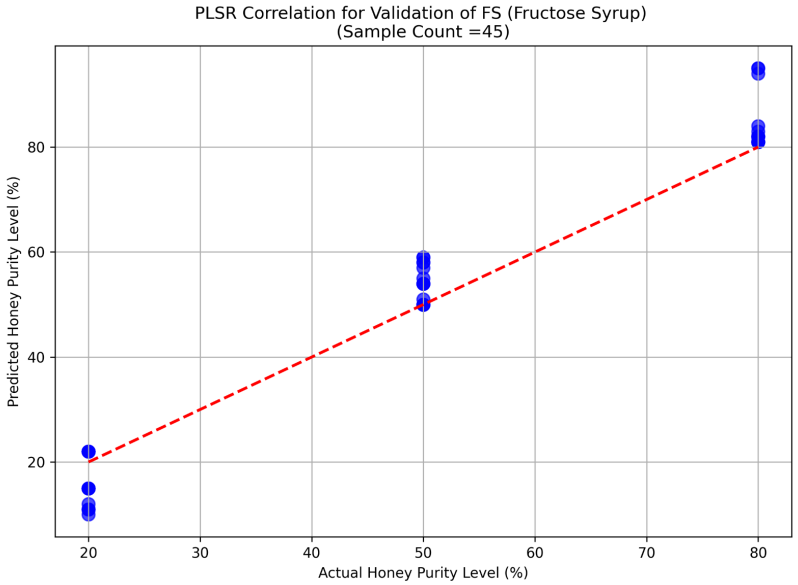
The visual alignment of predicted versus actual honey purity level shows reasonable clustering near the 1:1 diagonal, indicating consistent prediction trends even with fewer validation points of honey purity level as shown in Figure 12 which are correlated to each other. However, the minor prediction errors on RMSE are affected by using only 3 validation points, especially when outliers are present.



(a)



(b)



(c)

Fig. 12. Correlation between actual versus predicted sample that adulterated with (a) distilled water (DW), (b) apple cider vinegar (ACV), and (c) fructose syrup (FS)

To improve model robustness and generalizability, the future work should include a broader range of honey purity levels and additional independent sample sets.

3.4 Detection Limits and Analytical Performance

The section here evaluates detection capabilities, precision and repeatability, and performance of REVA In-Vitro system integrated with machine learning models for SBH adulteration and honey purity level analysis. This is to focus on determining the minimum honey purity levels that can be reliably detected, evaluating the repeatability of spectral measurements, and measuring the predictive performance of both classification and regression models.

3.4.1 Minimum detectable adulteration levels

The REVA In-Vitro spectrometer was evaluated across honey purity levels from 10% to 90% in 10% increments, demonstrating reliable detection and quantification for all three adulterant types tested. The lowest experimentally validated level was 10% honey purity (90% adulteration), where PLSR models achieved RMSE values of 0.27%, 0.36%, and 0.77% for DW, ACV, and FS adulterants respectively. The OVR classification model maintained 100% accuracy for adulterant type identification across the entire tested range. While distinct spectral signatures (Figures 1-3) suggest potential for lower concentration detection, actual detection limits below 10% remain undetermined without systematic evaluation.

Current results confirm reliable detection at 10% purity and below, representing commercially significant adulteration levels for fraud detection. Determination of absolute detection limits would require additional experimental validation with lower concentrations of adulterants.

3.4.2 Precision and repeatability

The experimental design incorporated 25 spectra per sample (5 scans × 5 replicates), enabling assessment of measurement of repeatability. PLSR training performance with $R^2 > 0.999$ and low RMSE values (0.27-0.77%) relative to the measurement range (10-100% purity) indicated good spectral consistency under controlled laboratory conditions.

Validation using independent samples maintained reasonable precision with R^2 values of 0.938-0.986, though RMSE increased to 2.90-6.10% as expected when applying models to new samples. The maintained linearity in actual versus predicted plots as in Figure 10 previously confirmed consistent measurement behavior across validation sets.

The observed repeatability applies specifically to the controlled laboratory conditions used, such as standardized sample preparation and consistent environmental factors. Assessment under varying operational conditions would require additional experimental validation.

3.4.3 Statistical performance metrics

The OVR classification model achieved 100% accuracy across training (n=700) and validation (n=50) datasets for the specific adulterant types and concentration ranges tested, corresponding to zero misclassification errors within the experimental framework. PLSR models demonstrated $R^2 > 0.999$ during training and 0.938-0.986 during validation, with RMSE values of 0.27-0.77% and 2.90-6.10% respectively. The higher validation RMSE for the FS model (6.10%) reflects spectral complexity from overlapping sugar signatures between fructose syrup and natural honey components, consistent with the challenging nature of sugar-based adulterant detection compared to water or vinegar additions.

These performance metrics apply to the controlled laboratory conditions and specific experimental parameters tested. Generalizability to different honey matrices, additional adulterant types, or varying environmental conditions requires systematic evaluation through broader validation studies.

4. Conclusions

This study demonstrated the effectiveness of Vis-NIR spectroscopy combined with OVR classification employed with PCA and PLSR models in detecting and quantifying adulteration in

stingless bee honey (SBH). The OVR model achieved 100% classification accuracy during training and validation, while PLSR models showed high predictive performance with R^2 values above 0.99 during training and above 0.93 during validation. These outcomes strengthen the validity of the models for deployment and future use in practical applications. Data expansion should be considered for model testing and validation by increasing the sample sizes for all classes to establish a more robust dataset, reducing variability, and increasing confidence in the model predictions. More samples are needed for model validation with the new honey variability in real-time. The use of preprocessed absorbance data preserved essential spectral features, enabling reliable differentiation across adulterants. Validation using new SBH samples confirmed the models' robustness and real-world applicability. These findings support the potential of the REVA In-Vitro platform for rapid, reagent-free honey authentication.

Acknowledgement

This research was funded by a grant from Ministry of Science, Technology and Innovation (MOSTI) of Malaysia (RMK-12 P5 – Platform Teknologi Fotonik (R.120005)).

References

[1] Wan, Chunfeng, Zhenwei Zhou, Siyuan Li, Youliang Ding, Zhao Xu, Zegang Yang, Yefei Xia, and Fangzhou Yin. "Development of a bridge management system based on the building information modeling technology." *Sustainability* 11, no. 17 (2019): 4583. doi: https://doi.org/10.3390/su11174583

[2] Kripka, Moacir, Victor Yepes, and Cleovir José Milani. "Selection of sustainable short-span bridge design in Brazil." *Sustainability* 11, no. 5 (2019): 1307. doi: https://doi.org/10.3390/su11051307

[3] Kaewunruen, Sakdirat, Jessada Sresakoolchai, and Zhihao Zhou. "Sustainability-based lifecycle management for bridge infrastructure using 6D BIM." *Sustainability* 12, no. 6 (2020): 2436. doi: https://doi.org/10.3390/su12062436

[4] Jensen, Jens Sandager. "Innovative and sustainable operation and maintenance of bridges." *Structure and Infrastructure Engineering* 16, no. 1 (2020): 72-83. doi: https://doi.org/10.1080/15732479.2019.1604772

[5] Kalajian, Khachig, Sonia Ahmed, and W. Youssef. "BIM in infrastructure projects." *International Journal of BIM and Engineering Science* 6, no. 2 (2023): 74-87. https://doi.org/10.54216/ijbes.060205

[6] Gorini, Davide Noè, Luigi Callisto, and Andrew J. Whittle. "Dominant responses of bridge abutments." *Soil Dynamics and Earthquake Engineering* 148 (2021): 106723. https://doi.org/10.1016/j.soildyn.2021.106723

[7] Gorini, Davide Noè, Luigi Callisto, and Andrew J. Whittle. "An inertial macroelement for bridge abutments." *Geotechnique* 72, no. 3 (2022): 247-259. https://doi.org/10.1680/jgeot.19.P.397

[8] Kharimah, Muhammad Ihsanul, and Benny Hidayat. "Pemodelan 3D Jembatan Lengkung, Studi Kasus Jembatan Bukit Sulap, Menggunakan Software Autodesk Revit." *Jurnal Bangunan, Konstruksi & Desain* 1, no. 3 (2023): 133-140. https://doi.org/10.25077/jbkd.1.3.133-140.2023

[9] Alizadehsalehi, Sepehr, Ahmad Hadavi, and Joseph Chuenhuei Huang. "From BIM to extended reality in AEC industry." *Automation in construction* 116, no. 103254 (2020): 1-13.https://doi.org/10.1016/j.autcon.2020.103254

[10] Parung, Herman, Muhammad Wihardi Tjaronge, Rudy Djamaluddin, Rita Irmawaty, Andi Arwin Amiruddin, Abdul Rahman Djamaluddin, Tri Harianto, Achmad Bakri Muhiddin, Ardi Arsyad, and Sitti Hijraini Nur. "Sosialisasi Aplikasi Teknologi Building Information Modelling (BIM) pada Sektor Konstruksi Indonesia." *Jurnal Tepat: Teknologi Terapan untuk Pengabdian Masyarakat* 2, no. 2 (2019): 112-119. https://doi.org/10.25042/jurnal_tepat.v2i2.82

[11] Calveen, Cayitho, and Andri Irfan Rifai. "Planning Analysis on Bridge Construction Possibilities and Challenges: A Review." *LEADER: Civil Engineering and Architecture Journal* 1, no. 2 (2023): 172-181. https://doi.org/10.37253/leader.v1i2.8098

[12] Firayanti, Dwiyana, and Ade Jaya Saputra. "Analysis Planning Upper Structure Bridge: A Case Barelang-2 Bridge, Batam." *LEADER: Civil Engineering and Architecture Journal* 1, no. 2 (2023): 91-102. https://doi.org/10.37253/leader.v1i2.7889

[13] Kadhim, Mohanad Mubdir, Ahmed Kareem Abdullah, Emad Kamil Hussein, and Faris Mohammed. "Experimental Investigation of Energy Harvesting by Employing Piezoelectric Element and Metallic Cantilever Beam." https://doi.org/10.37934/araset.53.1.186196

[14] Yatnikasari, Santi, Muhammad Noor Asnan, and Ulwiyah Wahdah Mufassirin Liana. "Alternatif Perencanaan Jembatan Rangka Baja Dengan Menggunakan Metode Lrfd Di Jembatan Gelatik Kota Samarinda." *Rang Teknik Journal* 4, no. 2 (2021): 282-294. https://doi.org/10.31869/rtj.v4i2.2518

[15] Yunus, Muhammad, and Zharin F. Syahdinar. "Stability Analysis of Aifa Bridge Abutment in Fafurwar District, Bintuni Bay Regency, West Papua Province." *EPI International Journal of Engineering* 2, no. 2 (2019): 162-171. https://doi.org/10.25042/epi-ije.082019.12

[16] Saputra, Ade Jaya. "An Evaluation of Improvement Plans Bridge Abutment Penaga in Bintan Island." *LEADER: Civil Engineering and Architecture Journal* 1, no. 2 (2023): 147-157. https://doi.org/10.37253/leader.v1i2.8079

[17] Rohmawati, Restu Farina, Paksitya Purnama Putra, and Indra Nurtjahjaningtyas. "Evaluasi Rancangan Abutment Jembatan Sungai Desa Kendalbulur Kecamatan Boyolangu Kabupaten Tulungagung." *Jurnal Teknik Sipil* 11, no. 1 (2022): 62-71.

[18] Yasin, Muhammad, Gusneli Yanti, and Shanti Wahyuni Megasari. "Analisis Abutment Jembatan Sei. Busuk Kabupaten Siak Sri Indrapura Provinsi Riau." *Siklus: Jurnal Teknik Sipil* 5, no. 1 (2019): 52-62. https://doi.org/10.31849/siklus.v5i1.2384

[19] Prasetyo, Sandika, Ester Priskasari, and Mohammad Erfan. "perencanaan Struktur Bawah (Abutment) Pada Pembangunan Jembatan Petak, Kabupaten Nganjuk." *Student Journal GELAGAR* 3, no. 1 (2021): 149-158. https://doi.org/10.55681/economina.v2i8.718

[20] Candra, Febrian, and Dwina Archenita. "Analysis Stability of Abutment on the Railway Bridge." *International Journal of Advanced Science Computing and Engineering* 4, no. 3 (2022): 233-242. https://doi.org/10.62527/ijasce.4.3.104

[21] Zabadi, Fairus. "ANALYSIS OF PILAR AND ABUTMENT FOUNDATIONS ON THE BATU RASANG-MAMBULU VILLAGE BRIDGE PROJECT TAMBELENGAN DISTRICT SAMPANG REGENCY." *Journal Innovation of Civil Engineering (JICE)* 3, no. 1 (2022): 1-11. https://doi.org/10.33474/jice.v3i1.15134

[22] Seong, Tan Hean, Teow Chien Xuan, and Shek Poi Ngian. "Parametric Study on Built-up Plate Girder to Eurocode 3 with Various Beam Span and Steel Grade." *Semarak Engineering Journal* 6, no. 1 (2024): 11-19.

[23] Fitriani, Heni, and Wina Prasetio Br Bangun. "Kesiapan Adopsi Building Information Modeling (Bim) Pada Konsultan Perencana Di Kota Palembang." *Teras Jurnal: Jurnal Teknik Sipil* 11, no. 2 (2021): 437-450. https://doi.org/10.37934/sej.6.1.1119

[24] Sidik, Dandi Muhamat, Nur Khotimah Handayani, and Fauzan Noer. "Analisis Perbandingan Volume Beton dan Besi Tulangan pada Struktur Gedung 10 Lantai di Kota Bandar Lampung antara Metode Konvensional dan Building Information Modeling (BIM) Autodesk Revit." In *Prosiding Seminar Nasional Teknik Sipil UMS*, pp. 591-597. 2023. https://doi.org/10.37729/suryabeton.v7i1.3031

[25] Hatmoko, Jati Utomo Dwi, Yulian Fundra, and Mochamad Agung Wibowo. "Investigating building information modelling (BIM) adoption in Indonesia construction industry." In *MATEC Web of Conferences*, vol. 258, p. 02006. EDP Sciences, 2019. https://doi.org/10.1051/matecconf/201925802006

[26] Zhou, Zijun. "An intelligent bridge management and maintenance model using BIM technology." *Mobile Information Systems* 2022, no. 1 (2022): 7130546. https://doi.org/10.1155/2022/7130546

[27] Puspita, Nurkhasanah Rina, and Fithriyah Patriotika. "BIM implementation in public construction projects in Indonesia." In *IOP Conference Series: Materials Science and Engineering*, vol. 1156, no. 1, p. 012008. IOP Publishing, 2021. https://doi.org/10.1088/1757-899x/1156/1/012008

[28] Sopaheluwakan, M. P., and T. J. W. Adi. "Adoption and implementation of building information modeling (BIM) by the government in the Indonesian construction industry." In *IOP Conference Series: Materials Science and Engineering*, vol. 930, no. 1, p. 012020. IOP Publishing, 2020. https://doi.org/10.1088/1757-899X/930/1/012020

[29] Costin, Aaron, Hanjin Hu, and Ronald Medlock. "Building information modeling for bridges and structures: Outcomes and lessons learned from the steel bridge industry." *Transportation Research Record* 2675, no. 11 (2021): 576-586. https://doi.org/10.1177/03611981211018691

[30] Bui, Nam. "Implementation of building information modeling in Vietnamese infrastructure construction: A case study of institutional influences on a bridge project." *The Electronic Journal of Information Systems in Developing Countries* 86, no. 4 (2020): e12128. https://doi.org/10.1002/isd2.12128

[31] Pantiga, Januar, and Anton Soekiman. "Kajian literatur implementasi building information modeling (BIM) di Indonesia." *Rekayasa Sipil* 15, no. 2 (2021): 104-110.https://doi.org/10.21776/ub.rekayasasipil.2021.015.02.4

[32] Hamzah, Amir, Selfin Anugrah Amdania, and Shcherbak Petr Nikolaevich. "Analysis of abutment safety factors against landslides on the Cipeundeuy bridge-Sukatani, Indonesia." *INTERNATIONAL JOURNAL ENGINEERING AND APPLIED TECHNOLOGY (IJEAT)* 4, no. 1 (2021): 1-10. https://doi.org/10.52005/ijeat.v4i1.46