Comparative Machine Learning Models for Classifying Sugar Syrup Adulteration in Honey using FTIR Spectroscopy

Deyla Prajna[1]*
[1]Department of Agricultural Technology, Jember State Polytechnic, Indonesia
[1]deylaprajna@polije.ac.id
*corresponding author

ABSTRACT

Honey adulteration with sugar syrups is a widespread issue that compromises product quality, safety, and consumer trust. This study used Fourier Transform Infrared (FTIR) spectroscopy in conjunction with machine learning methods to classify honey samples adulterated with sugar syrups at concentrations of 0%, 10%, 20%, and 40%. FTIR spectral data were pre-processed using Savitzky–Golay smoothing, followed by unsupervised and supervised analyses. Principal Component Analysis (PCA) revealed overlapping clusters among samples, indicating limited discriminatory power. In contrast, supervised classification models achieved higher classification accuracy, with Random Forest (RF) showing the best performance, achieving an accuracy of up to 100%. Followed by Linear Discriminant Analysis (LDA, 93.75%) and Support Vector Machine (SVM, 91.67%). These results demonstrate the strong potential of FTIR spectroscopy integrated with machine learning for rapid and non-destructive honey authentication. However, since the study utilized a publicly available dataset with limited sample information, future research should validate these models using larger, more diverse datasets to enhance reliability and applicability in real-world food authentication systems.

Keywords: Honey Adulteration; FTIR Spectroscopy; Machine Learning; Food Authentication

Article History		
Received : August, 29th 2025	Accepted: October, 28th 2025	Published: Nov, 7th 2025

I. INTRODUCTION

Honey adulteration is a worldwide issue that severely compromises product quality, consumer well-being, and overall market trust. Due to its high commercial value, nutritional benefits, and strong consumer demand, honey is frequently targeted by fraudulent practices aimed at economic gain. These practices include the direct addition of cheaper sweeteners, such as sugar syrups, corn syrups, high-fructose syrups, and industrial sugars, to increase volume and weight. Additionally, they involve misrepresenting the floral sources and geographical origins to attract premium pricing in niche markets. Such adulteration not only diminishes honey's intrinsic nutritional and therapeutic properties, including its antioxidant, antimicrobial, and wound-healing activities, but also poses potential health risks to consumers due to undisclosed ingredients and compromised purity standards [1], [2]. Moreover, the prevalence of adulterated honey erodes consumer confidence, disrupts fair trade, and negatively impacts genuine producers who adhere to quality standards, thereby necessitating the development of rapid, reliable, and cost-effective authentication techniques to safeguard food integrity and ensure public health.

The health consequences of honey adulteration are considerable, posing risks that extend beyond simple economic fraud to affect human well-being directly. Consumption of adulterated honey, often containing high levels of added sugar syrups or industrial sugars, can result in elevated blood glucose levels, contributing to metabolic disorders such as insulin resistance, obesity, and type 2 diabetes. Furthermore, prolonged intake of adulterated honey has been linked to potential organ damage, particularly affecting the liver and kidneys due to increased detoxification burden, as well as detrimental impacts on cardiovascular and neurological health, including heightened risks of heart disease and cognitive decline [3]. Alarmingly, sophisticated adulteration tactics employed by unscrupulous producers, such as ultrafiltration to remove detectable pollen and the use of syrups with similar spectral fingerprints, continue to elude routine quality assurance systems. This remains true despite the availability of detection tools that range from basic sensory and physicochemical analyses to advanced spectroscopic and chromatographic techniques [1]. The global prevalence of honey fraud is further exacerbated by inadequate traceability procedures in supply chains and the

DOI: http://dx.doi.org/10.25139/ijair.v7i1.10960

substantial financial incentives for adulteration, making it a pervasive issue that undermines food safety, authenticity, and fair market practices worldwide [2].

Conventional detection methods have been widely employed to identify adulterants and authenticate honey, primarily relying on techniques such as spectroscopy, gas chromatography (GC), and liquid chromatography (LC) [2]. Spectroscopy-based approaches, including UV-Vis, FTIR, and NIR, are often employed for rapid screening. In contrast, GC and LC coupled with mass spectrometry provide highly precise compositional profiling to detect specific adulterants or contaminants. Despite their analytical robustness, these techniques have limitations. They often require laborious and time-consuming sample preparation steps, specialized equipment, and trained personnel, which can restrict their application for routine quality control in many regions, particularly where laboratory infrastructure is limited. Moreover, the high operational costs and maintenance requirements of chromatographic systems pose additional barriers to widespread adoption. In response to these limitations, recent studies have investigated alternative and complementary methodologies aimed at improving detection efficiency, including the assessment of honey's physicochemical characteristics such as moisture content, electrical conductivity, diastase activity, and hydroxymethylfurfural (HMF) levels, as well as evaluations of antioxidant capacity and phenolic content [4]. These parameters not only provide insights into honey authenticity but also contribute to its nutritional and therapeutic evaluation, presenting a more accessible and cost-effective screening approach when combined with advanced statistical or chemometric analyses.

Advanced techniques have been developed to overcome the inherent limitations of conventional honey authentication methods, offering improved sensitivity, selectivity, and rapidity in detecting adulteration. One notable advancement is stable isotope ratio analysis, which examines isotopic signatures to differentiate between C3 and C4 plant-derived sugars. Specifically, the use of stable isotope analysis employing hexamethylenetetramine derived from honey monosaccharides has demonstrated enhanced sensitivity in identifying sugar syrup adulteration, providing a more precise determination of carbon isotope ratios [5]. This method improves upon standard isotope analysis by isolating the monosaccharide fraction, thus eliminating interference from proteins or other honey constituents that may obscure results. Additionally, High-Performance Liquid Chromatography with UV detection (HPLC-UV) integrated with chemometric analysis has emerged as an effective technique, enabling both qualitative classification and quantitative estimation of adulteration levels. Recent research reported that this method achieved complete classification accuracy, reliably identifying adulteration levels as low as 15%, thereby offering a robust and practical approach for routine monitoring [6]. Fluorescence spectroscopy has also evolved as a promising tool for authenticating honey due to its rapid, non-destructive nature and high sensitivity. Studies have shown that fluorescence spectroscopy can detect sugar syrup adulteration at levels of ≥10% by analyzing specific changes in emission spectra and frequency-doubled peak intensities, which are indicative of the presence of adulterants altering honey's intrinsic fluorescence characteristics [7]. Collectively, these advanced methods provide valuable alternatives or complements to conventional techniques, thereby enhancing the reliability and efficiency of honey adulteration detection, thereby protecting both consumer safety and market integrity.

Fourier Transform Infrared (FTIR) spectroscopy is recognized as a powerful analytical technique that has been extensively utilized in food authentication, particularly for detecting honey adulteration. This method operates based on the absorption of infrared radiation by chemical bonds within a molecule, causing atomic vibrations and rotations that produce distinct spectral fingerprints characteristic of specific functional groups and compounds [8]. FTIR spectroscopy enables rapid qualitative and quantitative analysis without the need for complex sample preparation, making it especially suitable for routine quality control in the food industry. In the context of honey authentication, FTIR has demonstrated considerable success in identifying and quantifying adulterants such as sugar syrups, glucose, and high-fructose corn syrup by analyzing characteristic absorption bands in the fingerprint region of the spectrum [4,9]. These spectral differences arise due to variations in carbohydrate structures and hydroxyl group vibrations between pure honey and adulterants.

Furthermore, the integration of FTIR spectral data with multivariate pre-processing and analytical approaches such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) enhances its discriminatory power. PCA reduces data dimensionality to reveal underlying patterns and groupings among samples, while HCA classifies them based on spectral similarity, facilitating clear differentiation between pure and adulterated honey samples. The non-destructive, rapid, and comprehensive nature of FTIR spectroscopy, combined with its high sensitivity and compatibility with chemometric tools, underscores its significant value in food quality management and authentication, combating fraudulent practices in the honey industry.

The integration of Fourier Transform Infrared (FTIR) spectroscopy with machine learning approaches has significantly advanced the capabilities of food authentication systems. FTIR spectroscopy provides rich spectral information reflecting the molecular composition of food samples. At the same time, machine learning algorithms can process and analyze these complex datasets to identify patterns, classify products, and detect adulteration with high precision. Among the most widely applied algorithms are Random Forest (RF), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA), each offering unique strengths in classification tasks. For instance, RF has been utilized to classify edible oils by detecting subtle compositional differences. At the same time, SVM has shown excellent performance in differentiating rice varieties based on their unique spectral fingerprints, and LDA has been effectively applied for classifying fruit juices according to authenticity and origin [10-12]. The success of these models is often enhanced by the application of spectral pre-processing techniques, such as Savitzky-Golay smoothing, which reduces noise and corrects baseline drift, and derivative transformations (first or second Derivative) that sharpen spectral features and resolve overlapping peaks [13]. These pre-processing steps are crucial for improving spectral resolution and amplifying subtle differences that are essential for accurate classification. The synergistic use of FTIR spectroscopy and machine

DOI: http://dx.doi.org/10.25139/ijair.v7i1.10960

learning thus provides a robust, rapid, and non-destructive analytical approach, contributing significantly to the development of effective food authentication and fraud prevention strategies in the food industry.

Particularly, when paired with FTIR spectral data, machine learning models have demonstrated outstanding efficiency in detecting honey adulteration and ensuring product authenticity. Numerous studies have demonstrated that applying machine learning to FTIR data improves classification performance by capturing the complex spectral variations associated with different adulteration levels. Research comparing various algorithms has indicated that Random Forest often outperforms other models due to its strong feature selection capabilities, ability to handle high-dimensional data without overfitting, and robust classification accuracy across diverse datasets [14]. Nevertheless, Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) have also delivered reliable results, with SVM effectively creating optimal hyperplanes for class separation and LDA excelling in cases where there are linear relationships between classes [15]. The comparative evaluation of these algorithms is crucial, as it provides insight into the suitability of models for specific data characteristics and analytical goals. These findings underscore the significance of comparison modeling techniques in selecting the most effective classification algorithm tailored to specific food authentication problems, such as honey adulteration detection, thereby enhancing the reliability and precision of food quality control systems.

Despite these technological advancements, a notable gap remains in the literature regarding the comparative evaluation of multiple machine learning models for classifying sugar syrup adulteration levels in honey based on FTIR spectral data. While individual algorithms such as Random Forest, Support Vector Machine, and Linear Discriminant Analysis have been successfully applied in food authentication contexts, limited research has systematically compared their performance specifically for honey adulteration detection using FTIR spectroscopy. Addressing this research gap, the present study aims to develop and evaluate the efficacy of these three supervised classification models in distinguishing honey samples adulterated with varying concentrations of sugar syrup. By evaluating the accuracy and robustness of each model across different pre-processing approaches, the study aims to identify the most suitable analytical strategy for practical application. The findings are anticipated to contribute to the development of rapid, reliable, and non-destructive techniques for honey authentication and quality assessment, thereby supporting industry needs in combating fraud and ensuring consumer protection.

II. METHOD

A. Materials

FTIR spectral data utilized in this study were sourced from the publicly available Kaggle dataset titled "FTIR Honey Dataset" [14], which can be accessed at *https://www.kaggle.com/datasets/alfaturk/ftir-honey-dataset*. This dataset comprises Fourier Transform Infrared (FTIR) spectra of honey samples intentionally adulterated with sugar syrup at varying concentrations, specifically at levels of 0%, 10%, 20%, and 40%. Each concentration level represents a distinct class, enabling evaluation of model classification performance across incremental adulteration. The dataset includes raw absorbance values across mid-infrared wavenumbers relevant to the chemical fingerprint region of honey. No additional sample preparation, spectral acquisition, or laboratory analysis was conducted, as this research employed the dataset in its provided format for computational modeling purposes. This approach facilitates a direct evaluation of machine learning models on available open-source data, thereby demonstrating the applicability of FTIR-based classification methods for real-world honey authentication scenarios.

B. Data Analysis

The overall workflow of this study is summarized in Fig.1, which outlines the sequential steps from FTIR spectral acquisition to data pre-processing, exploratory analysis, and supervised classification.

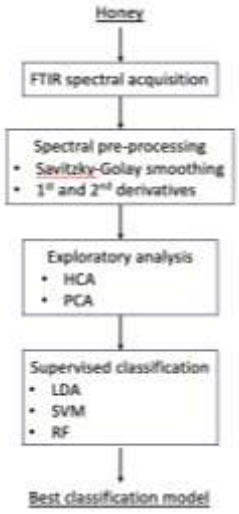

Fig.1. Workflow Of This Study

This structured approach ensures that each analytical stage, from signal enhancement to model evaluation, contributes systematically to identifying the most effective classification model for detecting honey adulteration. Spectral pre-processing was performed using the Savitzky-Golay (SG) smoothing technique, which included both first and second derivative transformations, to enhance spectral signal resolution and eliminate baseline variations that could interfere with classification accuracy. The SG method enhances peak sharpness and reduces noise, thus facilitating the extraction of meaningful spectral features essential for discriminative analysis. Following pre-processing, three supervised machine learning classification models using Random Forest (RF), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA). Constructed to predict the level of sugar syrup adulteration in honey samples. Each model was trained and tested to classify samples into their respective adulteration categories based on their FTIR spectra. The performance of these models was rigorously evaluated using classification accuracy as the primary metric, representing the proportion of correctly predicted samples across all classes. This methodological framework enables a comparative analysis of the models' effectiveness in addressing food authentication problems, particularly in detecting honey adulteration.

III. RESULT AND DISCUSSION

A. FTIR Spectral Characteristics

The FTIR spectra of honey samples with varying sugar syrup adulteration levels are illustrated in Fig.2. The unprocessed spectral profiles of the samples exhibit significant overlap, making visual differentiation between adulteration levels difficult due to the similar chemical compositions of honey and sugar syrups, both characterized by elevated amounts of carbohydrates that absorb in comparable spectral regions.

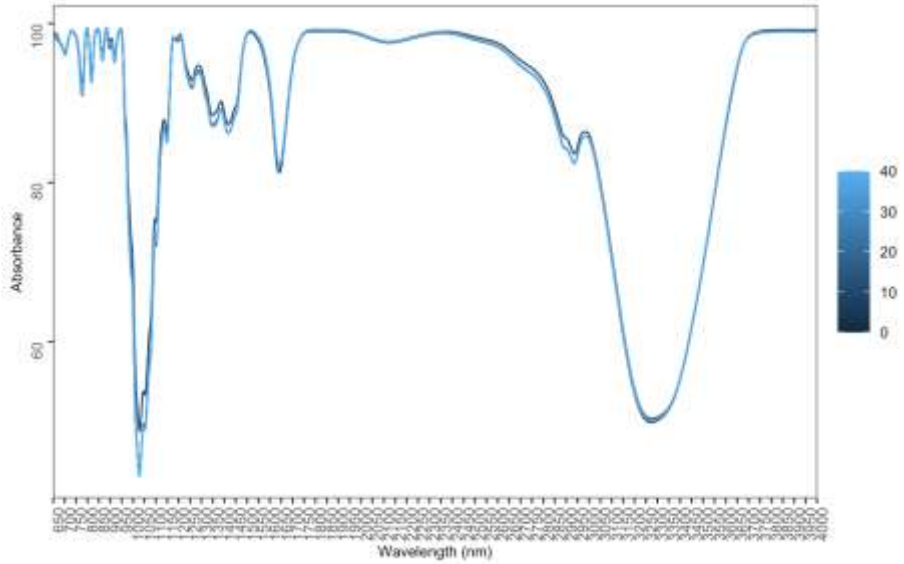

Fig.2. Spectra Of Honey Samples Without Pre-Treatment

The Savitzky–Golay (SG) filter was employed in this study to enhance spectral resolution and mitigate baseline fluctuations, thereby enabling more accurate analysis of the honey samples. This method effectively reduces noise and enhances peak features without significantly distorting the original spectral shape, making it suitable for spectroscopic studies, such as the analysis of complex biological matrices like honey, where overlapping peaks and baseline drifts often obscure meaningful chemical information [17,18]. Derivative pre-processing using SG filters markedly improves the quantification of adulteration levels by enhancing peak resolution. However, some studies have limitations in suppressing high-frequency noise and avoiding boundary artifacts [18]. Nonetheless, the continued application of SG pre-processing remains prevalent in FTIR-based food authentication research, as it offers a practical balance between noise reduction, spectral feature preservation, and computational efficiency — critical factors for ensuring reliable classification and quantification outcomes in spectroscopic analysis.

According to earlier research, Fourier Transform Infrared (FTIR) spectra of honey typically exhibit characteristic absorption in the carbohydrate fingerprint region, with $-CH_2$ and $-CH_3$ stretching vibrations around 2935-2885 cm^{-1} and broad bands attributed to water fractions at approximately 3333-1641 cm^{-1} [19]. Additional spectral features associated with carboxylic acids, amino acids, and carbohydrate ring vibrations have been identified as crucial indicators for the detailed characterization and authentication of honey samples [20]. However, it is often observed that they appear similarly in both pure and adulterated honey samples. This results in negligible variations across the spectra when assessed solely through visual inspection, underscoring the limitations of traditional qualitative analysis and highlighting the need for advanced multivariate or machine learning approaches to discriminate subtle differences attributable to adulteration effectively.

DOI: http://dx.doi.org/10.25139/ijair.v7i1.10960

B. Unsupervised Machine Learning (PCA)

Principal Component Analysis (PCA) was conducted on the pre-processed FTIR spectral data to investigate underlying patterns and potential distinctions among honey samples containing various levels of sugar syrup adulteration in Fig. 3. As a widely utilized multivariate statistical technique, PCA reduces the dimensionality of complex datasets by transforming the original correlated variables into a new set of uncorrelated variables known as principal components (PCs) that capture the most significant variance. The first principal component (PC1) accounts for the greatest variability, followed by the second (PC2), and so on, enabling efficient visualization and exploration of high-dimensional spectroscopic data. In this study, the application of PCA aimed to uncover clustering patterns or group separations that might correspond to different adulteration levels, thereby assessing whether unsupervised techniques could effectively discriminate between pure and adulterated honey samples. However, while PCA can reveal general variability trends within datasets, its capability to achieve distinct classification is often limited, necessitating the integration of supervised learning algorithms for accurate and predictive adulteration analysis.

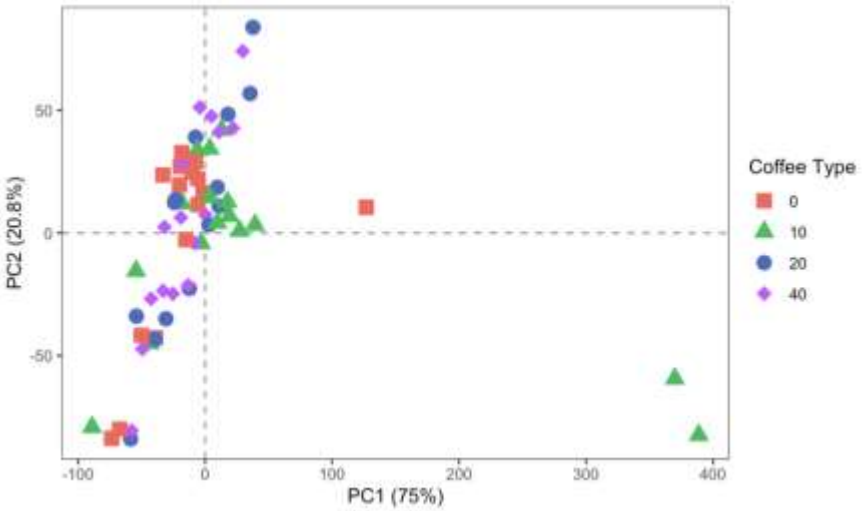

Fig.3. PCA of Honey Samples Without Pre-Treatment

This study found that principal component one (PC1) and principal component two (PC2) collectively accounted for approximately 95.8% of the cumulative variation in the FTIR spectral data, with PC1 contributing 75% and PC2 explaining an additional 20.8% of the total variability. Despite this high variance, the PCA scatter plot showed overlapping clusters, representing limited discrimination between sugar syrup adulteration levels. Thus, PCA effectively summarized the overall data variability, but it was insufficient in isolating distinctive spectral features, highlighting its limitation for classification tasks.

Previous research has documented similar challenges, wherein PCA successfully captured the majority of spectral variance yet failed to provide sufficient discrimination for accurate sample classification without supervised analysis [9, 23]. Although PCA serves as an effective tool for dimensionality reduction, data visualization, and exploratory analysis, PCA may lose subtle critical discriminative information because it emphasizes variance rather than classification boundaries [22]. Therefore, to achieve robust prediction models with enhanced discriminative capability, it becomes imperative to incorporate supervised classification methods such as Support Vector Machines (SVM), Random Forest (RF), or Linear Discriminant Analysis (LDA), which can capitalize on labelled data to effectively distinguish between pure and adulterated honey samples based on their intricate spectral differences.

C. Supervised Machine Learning (LDA, SVM, RF)

To accurately differentiate honey samples based on their levels of sugar syrup adulteration, supervised classification methods were applied following PCA analysis in Table I.

TABLE I
SUPERVISED RESULTS OF HONEY SAMPLES

		Raw Data	1st Derivative	2nd Derivative
LDA	No-Boruta	93.75%	58.33%	52.08%
	Boruta	-	72.92%	72.92%
SVM	No-Boruta	66.67%	91.67%	87.50%
	Boruta	-	64.58%	62.50%
RF	No-Boruta	100%	70.83%	100%
	Boruta	-	100%	100%

DOI: http://dx.doi.org/10.25139/ijair.v7i1.10960

Three widely recognized classification models were evaluated: Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Random Forest (RF). Each of these models offers distinct strengths, with LDA focusing on linear separation between classes, SVM providing robust performance for complex boundaries, and RF excelling in handling non-linear patterns and high-dimensional data. The models were assessed using spectral data under various pre-processing conditions: raw, first Derivative, and second Derivative, with and without Boruta feature selection. Model performance was evaluated using the accuracy metric, which reflects the proportion of correct predictions across all input data.

The Linear Discriminant Analysis (LDA) achieved its highest accuracy (93.75%) using raw data without Boruta, indicating that unprocessed spectral information preserved strong linear classification. Accuracy dropped to 58.33% for the first Derivative. Further, it declined to 52.08% for the second Derivative, both in the absence of Boruta, suggesting reduced linear separability. Nonetheless, the inclusion of Boruta with derivative-preprocessed data yielded a modest improvement in accuracy to 72.92% for both first and second derivatives. These findings indicate that LDA was most effective when applied directly to the raw spectral data, consistent with previous research findings that demonstrate reliability for FTIR-based honey classification, achieving up to 100% accuracy in honey powder analysis [23].

The SVM model attained its highest accuracy with first derivative data (91.67%), followed closely by second derivative data (87.50%), both without Bpruta. In contrast, raw data achieved lower accuracy (66.67%) without Boruta and fell below 65% with Boruta. These results demonstrate that SVM benefits significantly from derivative pre-processing, which has been shown to improve the performance of FTIR-based models in identifying syrup adulteration in honey samples [24]. Furthermore, SVM achieved 100% accuracy in detecting honey adulteration through the integration of fused near-infrared (NIR) and FTIR spectra, demonstrating its robust potential when combined with optimal pre-processing techniques [25].

The Random Forest (RF) model demonstrated exceptional performance, consistently achieving perfect accuracy (100%) when applied to both raw data and second derivative datasets, regardless of whether Boruta feature selection was used. Only the first derivative pre-processing without Boruta resulted in a 70.83% reduction in accuracy. These findings demonstrate that RF effectively captures complex, non-linear spectral relationships, providing robust classification without requiring extensive pre-processing. Similar studies have demonstrated RF's high accuracy in monofloral honey classification [26] and the discrimination of honey powder samples [23]. The 100% accuracy achieved here confirms RF's reliability for FTIR-based adulteration detection models, making them suitable for routine quality control and food fraud prevention applications.

Overall, RF outperformed both SVM and LDA, maintaining perfect accuracy under multiple conditions and showing exceptional robustness and adaptability to different data treatments. SVM ranked second, performing optimally with first derivative data, while LDA excelled with raw spectra but was sensitive to pre-processing transformations. Collectively, these results indicate that Random Forest is the superior classification model for detecting sugar syrup adulteration in honey using FTIR spectral data, offering unparalleled accuracy, stability, and operational simplicity for practical food authentication and fraud detection applications.

IV. CONCLUSION

This research demonstrated the efficacy of Fourier Transform Infrared (FTIR) spectroscopy, in conjunction with machine learning methods, as a reliable and non-destructive approach for detecting sugar syrup adulteration in honey. Among the tested models, Random Forest achieved the highest and most stable classification accuracy. At the same time, Support Vector Machine and Linear Discriminant Analysis performed effectively under specific pre-processing conditions, confirming the importance of data treatment in model optimization. These findings emphasize the potential of integrating spectroscopic data with artificial intelligence to enhance food authentication, quality control, and fraud prevention systems. Future research should incorporate larger and more diverse honey datasets, validate results with real-world samples, and explore the fusion of multimodal spectral data to develop robust and industry-ready AI tools for routine honey purity verification.

REFERENCES

[1] A. Girma Tura and D. Bersissa Seboka, "Review on Honey Adulteration and Detection of Adulterants in Honey," *International Journal of Gastroenterology*, vol. 4, no. 1, p. 1, 2020, doi: 10.11648/j.ijg.20200401.11.

[2] G. Zhang and W. Abdulla, "On honey authentication and adulterant detection techniques," *Food Control*, vol. 138, p. 108992, Aug. 2022, doi: 10.1016/j.foodcont.2022.108992.

[3] R. Fakhlaei *et al.*, "The Toxic Impact of Honey Adulteration: A Review," *Foods*, vol. 9, no. 11, p. 1538, Oct. 2020, doi: 10.3390/foods9111538.

[4] T. Damto, A. Zewdu, and T. Birhanu, "Impact of Different Adulterants on Honey Quality Properties and Evaluating Different Analytical Approaches for Adulteration Detection," *J Food Prot*, vol. 87, no. 4, p. 100241, Apr. 2024, doi: 10.1016/j.jfp.2024.100241.

[5] J. Li *et al.*, "The interaction between temperature and citric acid treatment in the anaerobic fermentation of Chinese cabbage waste," *J Clean Prod*, vol. 383, p. 135502, Jan. 2023, doi: 10.1016/j.jclepro.2022.135502.

[6] C. Egido, J. Saurina, S. Sentellas, and O. Núñez, "Honey fraud detection based on sugar syrup adulterations by HPLC-UV fingerprinting and chemometrics," *Food Chem*, vol. 436, p. 137758, Mar. 2024, doi: 10.1016/j.foodchem.2023.137758.

[7] S. Yan, M. Sun, X. Wang, J. Shan, and X. Xue, "A Novel, Rapid Screening Technique for Sugar Syrup Adulteration in Honey Using Fluorescence Spectroscopy," *Foods*, vol. 11, no. 15, p. 2316, Aug. 2022, doi: 10.3390/foods11152316.
[8] Y. Gong, X. Chen, and W. Wu, "Application of fourier transform infrared (FTIR) spectroscopy in sample preparation: Material characterization and mechanism investigation," *Advances in Sample Preparation*, vol. 11, p. 100122, Aug. 2024, doi: 10.1016/j.sampre.2024.100122.
[9] O. Elhamdaoui, A. El Orche, A. Cheikh, B. Mojemmi, R. Nejjari, and M. Bouatia, "Development of Fast Analytical Method for the Detection and Quantification of Honey Adulteration Using Vibrational Spectroscopy and Chemometrics Tools," *J Anal Methods Chem*, vol. 2020, pp. 1–9, Dec. 2020, doi: 10.1155/2020/8816249.
[10] P. Tsakanikas, A. Karnavas, E. Z. Panagou, and G.-J. Nychas, "A machine learning workflow for raw food spectroscopic classification in a future industry," *Sci Rep*, vol. 10, no. 1, p. 11212, Jul. 2020, doi: 10.1038/s41598-020-68156-2.
[11] J. L. P. Calle, M. Ferreiro-González, A. Ruiz-Rodríguez, D. Fernández, and M. Palma, "Detection of Adulterations in Fruit Juices Using Machine Learning Methods over FT-IR Spectroscopic Data," *Agronomy*, vol. 12, no. 3, Mar. 2022, doi: 10.3390/agronomy12030683.
[12] D. Prajna *et al.*, "Enhanced Differentiation of Wild and Feeding Civet Coffee Using Near-Infrared Spectroscopy with Various Sample Pretreatments and Chemometric Approaches," *Horticulturae*, vol. 9, no. 7, Jul. 2023, doi: 10.3390/horticulturae9070778.
[13] D. Prajna, M. Barea-Sepúlveda, J. L. P. Calle, D. Suhandy, W. Setyaningsih, and M. Palma, "Headspace gas chromatography with various sample preparation and chemometric approaches to improve discrimination of wild and feeding civet coffee," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2024. doi: 10.1088/1755-1315/1356/1/012037.
[14] Alfaturk, "FTIR Honey Dataset," Kaggle. https://www.kaggle.com/datasets/alfaturk/ftir-honey-dataset.
[15] D. K. Agustika, M. R. Nawawi, R. Prasetyowati, S. H. Hidayat, D. D. Iliescu, and M. S. Leeson, "Savitzky-Golay Parameter Optimization by using Linear Discriminant Analysis for FTIR Spectra," in *2022 IEEE 7th Forum on Research and Technologies for Society and Industry Innovation (RTSI)*, IEEE, Aug. 2022, pp. 123–128. doi: 10.1109/RTSI55261.2022.9905171.
[16] A. S. Tsagkaris *et al.*, "Investigating the impact of spectral data pre-processing to assess honey botanical origin through Fourier transform infrared spectroscopy (FTIR)," *Journal of Food Composition and Analysis*, vol. 119, p. 105276, Jun. 2023, doi: 10.1016/j.jfca.2023.105276.
[17] G. G. Dumancas and H. Ellis, "Comprehensive examination and comparison of machine learning techniques for the quantitative determination of adulterants in honey using Fourier infrared spectroscopy with attenuated total reflectance accessory," *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 276, p. 121186, Aug. 2022, doi: 10.1016/j.saa.2022.121186.
[18] M. Schmid, D. Rath, and U. Diebold, "Why and How Savitzky–Golay Filters Should Be Replaced," *ACS Measurement Science Au*, vol. 2, no. 2, pp. 185–196, Apr. 2022, doi: 10.1021/acsmeasuresciau.1c00054.
[19] F. Guyon, E. Logodin, D. A. Magdas, and L. Gaillard, "Potential of FTIR- ATR diamond in discriminating geographical and botanical origins of honeys from France and Romania," *Talanta Open*, vol. 3, p. 100022, Aug. 2021, doi: 10.1016/j.talo.2020.100022.
[20] A. G. Ioannou *et al.*, "Highlighting the Potential of Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) Spectroscopy to Characterize Honey Samples with Principal Component Analysis (PCA)," *Anal Lett*, vol. 56, no. 5, pp. 789–806, Mar. 2023, doi: 10.1080/00032719.2022.2103143.
[21] N. S. binti Zulkiflee, A. S. Zoolfakar, R. Abdul Rani, D. Aryani, and M. Zolkapli, "Detection and Classification of Honey Adulteration Combined with Multivariate Analysis," *International Journal of Integrated Engineering*, vol. 14, no. 3, Jun. 2022, doi: 10.30880/ijie.2022.14.03.028.
[22] J. Zheng and C. Rakovski, "On the Application of Principal Component Analysis to Classification Problems," *Data Sci J*, vol. 20, Aug. 2021, doi: 10.5334/dsj-2021-026.
[23] A. Matwijczuk *et al.*, "Classification of Honey Powder Composition by FTIR Spectroscopy Coupled with Chemometric Analysis," *Molecules*, vol. 27, no. 12, p. 3800, Jun. 2022, doi: 10.3390/molecules27123800.
[24] G. Dumancas, H. Ellis, J. Neumann, and K. Smith, "Comparison of Various Signal Processing Techniques and Spectral Regions for the Direct Determination of Syrup Adulterants in Honey Using Fourier Transform Infrared Spectroscopy and Chemometrics," *Chemosensors*, vol. 10, no. 2, p. 51, Jan. 2022, doi: 10.3390/chemosensors10020051.
[25] L. F. Huang *et al.*, "Fingerprint developing of coffee flavor by gas chromatography-mass spectrometry and combined chemometrics methods," *Anal Chim Acta*, vol. 588, no. 2, pp. 216–223, Apr. 2007, doi: 10.1016/j.aca.2007.02.013.
[26] M. Ciulu *et al.*, "Classification of Unifloral Honeys from SARDINIA (Italy) by ATR-FTIR Spectroscopy and Random Forest," *Molecules*, vol. 26, no. 1, p. 88, Dec. 2020, doi: 10.3390/molecules26010088.