Honey Adulteration Detection through Hyperspectral Imaging and Machine Learning

Hazrat Usman[1], Anna Amjad[1], Maryam Mahsal Khan[1], Sumayyea Salahuddin[2]
[1]Department of Computer Science (CECOS University of IT and Emerging Sciences, Peshawar, Pakistan).
[2]Department of Computer Systems Engineering (University of Engineering and Technology, Peshawar, Pakistan).
*Correspondence: Maryam Mahsal Khan, maryam.khan@cecos.edu.pk,

Citation | Usman. H, Amjad. A, Khan. M. M, Salahuddin. S, "Honey Adulteration Detection through Hyperspectral Imaging and Machine Learning", IJIST, Special Issue pp 30-36, May 2024

Received | May 02, 2024, **Revised |** May 09, 2024, **Accepted |** May 13, 2024, **Published |** May 20, 2024.

Introduction/Importance of Study:

The purity and authenticity of honey are paramount for ensuring consumer trust and maintaining the integrity of the honey industry. There is a pressing need for advanced and efficient detection methods to increase the prevalence of honey adulteration.

Novelty statement:

Our research provides a solution to the challenge of predicting the change in adulterated honey properties through hyperspectral imaging and advanced machine learning algorithms, filling a critical gap in existing methodologies.

Material and Method:

A publicly available dataset with spectral features, extracted through hyperspectral imaging, across different classes of honey and adulteration levels has been examined and various machine learning models were developed to identify honey adulteration concentration and type of honey. The dataset was balanced and a five-fold cross-validation technique was used to train the machine learning models.

Result and Discussion:

Random forest was found to perform better in three identified scenarios i.e. (a) type of honey (b) adulteration level (c) both (a, b); with a maximum average accuracy of 99.69% performing better than the one reported in the literature (95%). For both single-output and multiple-output ML models, the trend in feature importance was observed. The single model identifying the class of honey utilized low and mid-frequency spectra while the multi-model used mid-frequency spectrum only.

Concluding Remarks:

The proposed approach aims to provide an accurate and cost-effective solution to address the challenges associated with honey adulteration, contributing to the enhancement of honey quality assessment and consumer confidence.

Keywords: Honey Fraud Detection; Hyperspectral imaging; Machine Learning; Random Forest; XG Boost.

Introduction:

Honey has been since ancient times valued for its unique flavors and nutritional qualities. Honey production is a profitable market. Due to the increase in demand for honey because of population growth, honey producers are tempted to commit fraud. Pure honey is diluted with common adulterants, like sugar syrups and other sweets. The honey contamination results not only in quality compromise but also severe health issues. Conventional approaches to honey adulteration are unable to identify the various adulterants in the honey. This emphasizes the need to develop and design sophisticated techniques that guarantee quick and accurate findings. In this setting, hyperspectral imaging becomes an effective tool that provides a non-destructive way to obtain detailed spectral information from honey samples over a wide range of wavelengths. With the use of this technology, complex chemical fingerprints may be extracted, making it possible to distinguish between genuine and fake honey. However, advanced analytical tools like machine learning are needed to fully utilize hyperspectral data. Research in this area is being done by the current study.

This work outlines a methodical approach that includes gathering hyperspectral data from honey samples, optimizing data quality through preprocessing, and utilizing multiple machine learning algorithms to achieve robust detection. One of the outcomes of the research is a flexible and all-encompassing strategy that can protect the integrity of the honey business. The goal of the suggested method is to combine machine learning and hyperspectral imaging to transform the detection of adulterated honey. By combining these technologies, this study aims to provide a smart and effective solution that will guarantee the production and consumption of unadulterated honey for years to come. The current study aims to develop an intelligent system that can detect adulteration in honey using machine learning algorithms across three scenarios.

- Identify the class or type of honey.
- Identify the level of adulteration of sugar syrup, in the honey class.
- Identify both the type of honey and the level of adulteration in one go.

Literature Review:

Existing Quality Assurance Techniques:

Most honey's plant sources are categorized chemically; however, more conventional methods still include honey specialists tasting and smelling the honey. Pollen analysis and assays for specific components that make up different types of honey are among the chemical measures [1]. Numerous techniques have been put up to identify the adulteration of honey with sugar. High-pressure liquid chromatography (HPLC) [2], deuterium nuclear magnetic resonance (NMR) spectroscopy [3][4], mass spectroscopy of the carbon isotope ratio [5][6], and FTIR spectroscopy [7][8] can be used to detect adulterated honey.

The detection of honey adulteration with cane sugar using Fourier transform infrared (FTIR) spectroscopy has been the subject of several studies [9]. These studies assessed adulteration in cane sugar concentrations ranging from 0.5 to 25%. In [10], a single variety of honey was utilized to estimate the sugar concentration with an accuracy of 93.75% utilizing statistical techniques and artificial neural networks. When three different varieties of honey were used to classify adulteration, the classification accuracy was less than 80% [9]. These studies demonstrate that it is feasible to anticipate adulteration in honey by combining spectroscopic and machine learning approaches; yet, the capacity to forecast sugar content across a variety of honey varieties has to be enhanced.

By extending spectroscopy and enabling the use of spatial information in addition to spectral information, hyperspectral imaging is a potential method for ensuring the quality of food [11]. Instead of only capturing the spectrum at one spot on the item, spatial information enables the image to highlight certain flaws like bruising on fruit at a specific area [12]. Numerous food quality applications, such as those involving meat, fish, fruit, vegetables, and cereals, have made use of hyperspectral imaging [5].

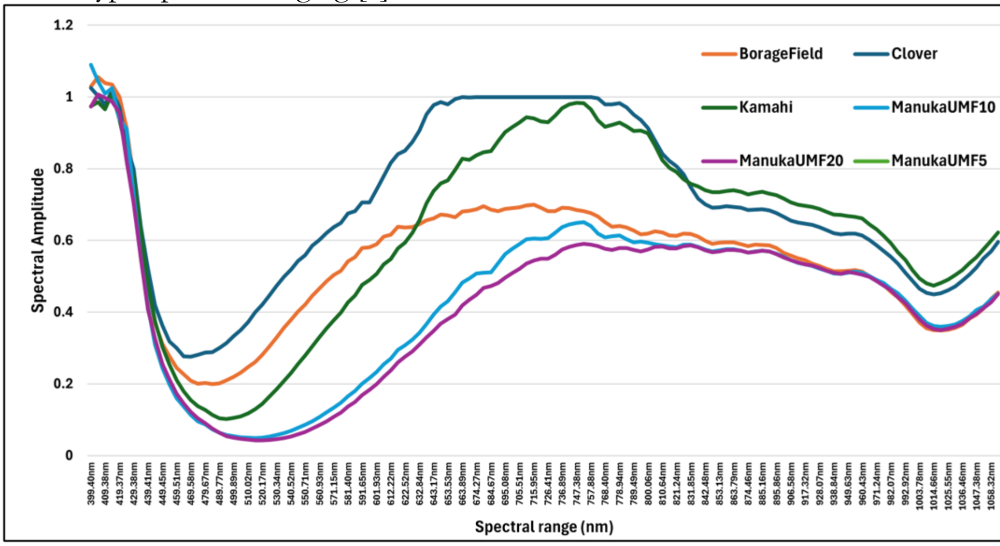

Figure 1. Hyperspectral response of 6 different types of honey classes with an adulteration concentration level of 50%.

A hyperspectral imaging technique has been developed to determine the botanical source of honey [13][14][15][16]. With 90% accuracy, the botanical ancestry of 21 different types of honey was predicted [16]. These techniques used a class embodiment autoencoder (CEAE) and support vector machines (SVM) to classify the data, which was obtained via a hyperspectral imaging system as detailed in [17]. In [13], 52 samples from five distinct types of honey were identified botanically with a 90% classification accuracy in a small data set.

Adulteration Dataset:

The dataset [18] comprises 12 unique honey products sourced from seven different brands and characterized by 11 botanical origin labels. Six independent samples were collected for each type of honey, with an equal distribution between Manuka honey, a high-quality honey variant from New Zealand, and other types of New Zealand honey. Throughout the dataset creation process, images of all honey varieties were captured at varying sugar concentrations (5, 10, 25, 50). The detailed composition of the dataset is presented in Table 1. Figure 1 shows the hyperspectral response of six different types of honey classes with a honey adulteration level of 50%. The drift in the spectrum indicates that the effect of adulteration on the spectral response of honey is different and non-linear, where AI would aid in interpreting the trends accordingly.

Material and Methods:

Table 1: The overall makeup of the adulterated honey data set from each brand and botanical origins label of honey. taken from [18]

Class	**Adulteration Concentration**					
	0%	5%	10%	25%	50%	Sum
Clover	150	150	300	300	300	1200
Multi Floral	150	150			150	450
ManukaUMF5		150	150	150	150	600
ManukaUMF15		150	150	150	150	600
ManukaUMF20		150	150	150	150	600
ManukaUMF10		150	150	150	125	575
Manuka Blend		150		150	150	450
Borage Field	150	150	150	150	150	750
Kamahi	150	150	150	150	150	750

Rewarewa	150	150	150			450
Manuka Blend	150	150	150	150	150	750
Manuka	150	300	300	300	300	1350

Figure 2 shows the methodology of the current study in which we have focused on the detection of the following three scenarios. Given a honey sample,

- Task A: To identify the class of honey.
- Task B: To identify the adulteration concentration level.
- Task C: To identify both the class and adulteration concentration level.

The representation of samples is not balanced in all these tasks. In order to reduce bias in the generated machine learning (ML) models, we have then employed balancing the dataset via a popular and common technique called Synthetic Minority Oversampling Technique (SMOTE) [19]. In this manner, accurate measurement of ML performance metrics can be recorded accordingly.

- TaskA represents 12 Honey classes (first column in Table 1). After balancing, each class represented 1350 samples.
- For TaskB, honey samples consist of adulteration concentration levels of 0%, 5%, 10%, 25%, and 50% in the dataset. After balancing the set a total of 1950 samples per class were generated.
- Finally, Task C is to identify honey type and adulteration concentration level while using one ML model. After balancing 300 samples per class were generated.

The dataset created in this manner for all three tasks is then forwarded to the ML model for creation and evaluation. A five-fold cross-validation was used to train and test the ML models. From Figure 3, given a dataset, three different representations of the dataset are generated that are then balanced by using SMOTE. Machine learning algorithms in particular Random Forest (RF), Support Vector Machine (SVM), and Extra Gradient boosting Trees (XG Boost) were used in generating the models. The performance of the algorithms across the different tasks is reported in Table II-IV.

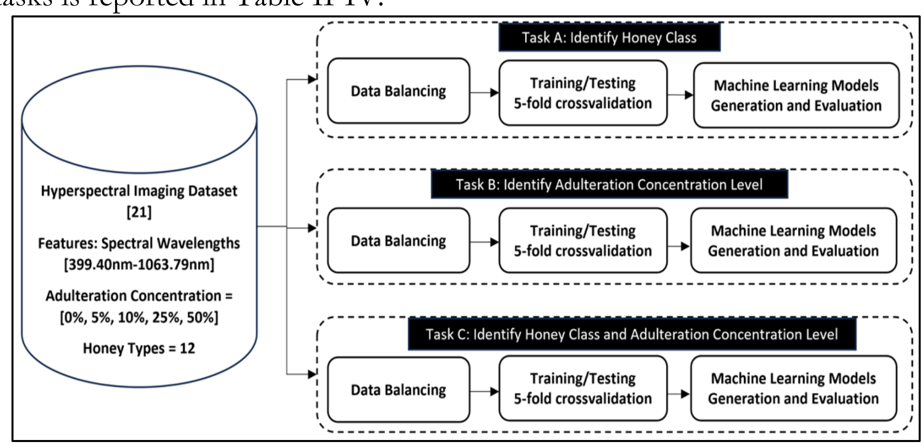

Figure 2: Methodology and execution of various tasks in the current research study.

Result and Discussion:

Table 2-4 shows the performance of the ML models across three different tasks (Task A, Task B, and Task C) shown in Figure 2. In order to access the performance of the ML algorithm, various performance metrics are used. In the current study ML performance metrics namely accuracy, precision, recall, and f1-score are reported herewith.

In Table 2, for Task A, out of the three ML models, the RF model performed best with an accuracy of 99.89% while SVM scored an accuracy of 85.35%. In this task, each honey type whether pure or adulterated is categorized as one. The result indicates that the spectral parameters derived from hyperspectral imaging can identify the type of honey investigated.

Table 2: Performance of ML Models on task A, given a honey sample, identify the Type of Honey.

ML Model	Accuracy	Precision	Recall	F1 Score
RF	**0.99899**	0.999043	0.998991	0.999013
SVM	0.853535	0.864275	0.854573	0.851682
XG Boost	0.997306	0.997337	0.997351	0.997334

Table 3: Performance of ML Models on task B, given a honey sample, identify the concentration adulteration Level

ML Model	Accuracy	Precision	Recall	F1 Score
RF	0.996923	0.99691	0.996953	0.996926
SVM	0.522564	0.51252	0.526156	0.4714
XG Boost	0.995897	0.995878	0.995958	0.995908

While in Table 3, Task B, RF performed best with an accuracy of 99.69%. SVM was found to perform with an accuracy of 52.25%. In this task, the different types of honey were grouped based on their adulteration concentration level. The adulterant used in the dataset was sugar syrup. The task was particularly challenging as honey also contains natural sugar.

Table 4: Performance of ML Models on task C, given a Honey sample, identify both class and type of Honey.

ML Model	Accuracy	Precision	Recall	F1 Score
RF	0.998936	0.998981	0.998897	0.99893
SVM	0.818794	0.84086	0.821589	0.800515
XG Boost	0.992553	0.992895	0.992604	0.992638

In Table 4, Task C, RF performed with an accuracy of 99.92%. The identification of honey and adulteration levels via a single ML model was found to be better than two distinct models. This might be due to the reason honey types represent unique spectral signatures. Hence the change in signature on the spectral properties is reflected differently among different honey types.

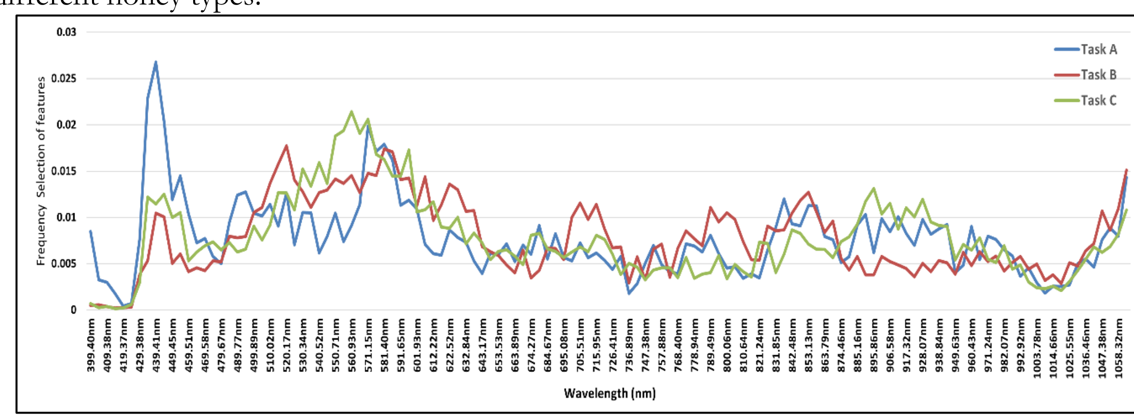

Figure 3: Feature significance graph of the RF models generated for Task A, Task B, and Task C respectively.

Figure 3 shows the feature significance graph obtained from the RF models of the three tasks A, B, and C. The feature importance shows that for Task A two identifiable peaks are obtained at a spectral range of 439.41nm and 566.04nm, for Task B around 520.17nm and 586nm while for Task C spectral range of 560nm and 596.79nm were found to be the important spectral features in these tasks. It's interesting to note that from Task A to Task C a shift in peak features is observed.

Table 5, shows the comparison of the ML models with the ones reported in literature [13] on the dataset. As the dataset was not balanced in the reported literature hence f1-score is low. Due to the balancing of the dataset, our study was able to improve on these metrics.

Table 5: Comparison of ML models with others reported in the literature

Task	Accuracy	Precision	Recall	F1 Score
Adulteration Concentration	0.996923	0.99691	0.996953	0.996926
Adulteration Concentration [18]	0.951	x	x	0.940

Conclusion:

The adulteration of honey has become a widespread practice aimed at increasing economic benefits, however, it has been shown to have detrimental effects on an individual's health. The current study explores the potential of machine learning algorithms in the accurate identification of honey adulteration on a recently publicly available dataset. SMOTE algorithm was used to balance the dataset. Random Forest, Support Vector Machine, and XGBoost algorithms were used to generate models whereby RF was found to perform better than the other two algorithms in the identification of the quality of honey. In comparison to the reported study (95%), our study produced an accuracy of 99.69% on the same dataset. This indicates the potential of ML algorithms in the accurate identification and quantification of honey adulteration.

Acknowledgment: We acknowledge the Department of ORIC at CECOS University of IT & Emerging Sciences for funding resources.

Author's Contribution: All authors have equally contributed to the paper.

Conflict of Interest: The authors have no conflicts of interest to declare.

Project Details: NA

References:

[1] A. Noviyanto, W. Abdullah, W. Yu, and Z. Salcic, "Research trends in optical spectrum for honey analysis," 2015 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA ASC 2015, pp. 416–425, Feb. 2016, doi: 10.1109/APSIPA.2015.7415305.

[2] C. Cordella, I. Moussa, A. C. Martel, N. Sbirrazzouli, and L. Lizzani-Cuvelier, "Recent Developments in Food Characterization and Adulteration Detection: Technique-Oriented Perspectives," J. Agric. Food Chem., vol. 50, no. 7, pp. 1751–1764, Mar. 2002, doi: 10.1021/JF011096Z.

[3] P. Lindner, E. Bermann, and B. Gamarnik, "Characterization of Citrus Honey by Deuterium NMR," J. Agric. Food Chem., vol. 44, no. 1, pp. 139–140, 1996, doi: 10.1021/JF950359K.

[4] S. Giraudon, M. Danzart, and M. H. Merle, "Deuterium Nuclear Magnetic Resonance Spectroscopy and Stable Carbon Isotope Ratio Analysis/Mass Spectrometry of Certain Monofloral Honeys," J. AOAC Int., vol. 83, no. 6, pp. 1401–1409, Nov. 2000, doi: 10.1093/JAOAC/83.6.1401.

[5] S. T. Brookes, A. Barrie, and J. E. Davies, "A Rapid 13C/12C Test for Determination of Corn Syrups in Honey," J. AOAC Int., vol. 74, no. 4, pp. 627–629, Jul. 1991, doi: 10.1093/JAOAC/74.4.627.

[6] J. W. White, K. Winters, P. Martin, and A. Rossmann, "Stable Carbon Isotope Ratio Analysis of Honey: Validation of Internal Standard Procedure for Worldwide Application," J. AOAC Int., vol. 81, no. 3, pp. 610–619, May 1998, doi: 10.1093/JAOAC/81.3.610.

[7] S. Gok, M. Severcan, E. Goormaghtigh, I. Kandemir, and F. Severcan, "Differentiation of Anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate analysis," Food Chem., vol. 170, pp. 234–240, Mar. 2015, doi: 10.1016/J.FOODCHEM.2014.08.040.

[8] S. Sivakesava and J. Irudayaraj, "Detection of inverted beet sugar adulteration of honey by FTIR spectroscopy," J. Sci. Food Agric., vol. 81, no. 8, pp. 683–690, Jun. 2001, doi: 10.1002/JSFA.858.

[9] S. Sivakesave and J. Irudayaraj, "Prediction of Inverted Cane Sugar Adulteration of

Honey by Fourier Transform Infrared Spectroscopy," J. Food Sci., vol. 66, no. 7, pp. 972–978, Sep. 2001, doi: 10.1111/J.1365-2621.2001.TB08221.X.
[10] J. Irudayaraj, F. Xu, and J. Tewari, "Rapid Determination of Invert Cane Sugar Adulteration in Honey Using FTIR Spectroscopy and Multivariate Analysis," J. Food Sci., vol. 68, no. 6, pp. 2040–2045, Aug. 2003, doi: 10.1111/J.1365-2621.2003.TB07015.X.
[11] G. ElMasry and D. W. Sun, "Principles of Hyperspectral Imaging Technology," Hyperspectral Imaging Food Qual. Anal. Control, pp. 3–43, Jan. 2010, doi: 10.1016/B978-0-12-374753-2.10001-2.
[12] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging – an emerging process analytical tool for food quality and safety control," Trends Food Sci. Technol., vol. 18, no. 12, pp. 590–598, Dec. 2007, doi: 10.1016/J.TIFS.2007.06.001.
[13] S. Minaei et al., "VIS/NIR imaging application for honey floral origin determination," Infrared Phys. Technol., vol. 86, pp. 218–225, Nov. 2017, doi: 10.1016/J.INFRARED.2017.09.001.
[14] A. Noviyanto and W. H. Abdulla, "Honey botanical origin classification using hyperspectral imaging and machine learning," J. Food Eng., vol. 265, p. 109684, Jan. 2020, doi: 10.1016/J.JFOODENG.2019.109684.
[15] A. Noviyanto and W. H. Abdulla, "Signifying the information carrying bands of hyperspectral imaging for honey botanical origin classification," J. Food Eng., vol. 292, p. 110281, Mar. 2021, doi: 10.1016/J.JFOODENG.2020.110281.
[16] T. Phillips and W. Abdulla, "Class Embodiment Autoencoder (CEAE) for classifying the botanical origins of honey," Int. Conf. Image Vis. Comput. New Zeal., vol. 2019-December, Dec. 2019, doi: 10.1109/IVCNZ48456.2019.8961004.
[17] A. Noviyanto and W. H. Abdulla, "Honey dataset standard using hyperspectral imaging for machine learning problems," 25th Eur. Signal Process. Conf. EUSIPCO 2017, vol. 2017-January, pp. 473–477, Oct. 2017, doi: 10.23919/EUSIPCO.2017.8081252.
[18] T. Phillips and W. Abdulla, "A new honey adulteration detection approach using hyperspectral imaging and machine learning," Eur. Food Res. Technol., vol. 249, no. 2, pp. 259–272, Feb. 2023, doi: 10.1007/S00217-022-04113-9/TABLES/6.
[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/JAIR.953.