International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Fast and Robust Quality Assessment of Honey Using Near Infrared Spectroscopy

P. Sudeep Sai[1], G. Sailesh Varma[2], Dr. G. Rajalakshmi[3]

[1,2]Department of Electronics and Communication Engineering Student of Sathyabama Institute of Science and technology Chennai, India
[3]M.E.Ph. D, Department of Electronics and Communication Engineering Associate Professor of Sathyabama Institute of Science and Technology Chennai, India

Abstract

Assessing the quality of honey is crucial to upholding industry standards and guaranteeing customer safety. Using important physicochemical characteristics like pH, viscosity, electrical conductivity, glucose, fructose, and pollen analysis, this study uses machine learning approaches to forecast honey purity and pricing. To assess the predictive power of the RF and Gradient Boosting Regressor models, we put them into practice and compare them. To improve model performance, the dataset is preprocessed using techniques like feature scaling and categorical encoding. Gradient Boosting outperforms RF in terms of prediction accuracy, as evidenced by evaluation criteria like R2 score. According to the findings, machine learning can help with honey quality evaluation and verification in an efficient manner, offering producers, consumers, and regulatory agencies useful information.

Keywords: Honey,Adulteration,quality,NIR, preprocessing, validation

1. INTRODUCTION

Due to its nutritional and therapeutic qualities, honey is a popular natural sweetener. However, questions about honey's quality, purity, and market pricing have been highlighted by the growing incidence of adulteration. Conventional techniques for evaluating the quality of honey depend on microscopic and chemical analysis, which can be laborious and need professional oversight. In order to guarantee the authenticity of honey, automated and data-driven methods are becoming more and more necessary. Utilising physicochemical characteristics including pH, viscosity, glucose, and electrical conductivity, machine learning approaches in particular, regression-based models offer a potential way to forecast the price and purity of honey.

Based on a dataset that includes important honey attributes, we use RF and Gradient Boosting Regressor models in this work to forecast honey purity and pricing. To enhance model performance, the dataset is preprocessed using methods like feature scaling and categorical encoding. We hope to ascertain the best method for honey quality analysis by contrasting the predicted accuracy of these two models. The study offers insights into the primary determinants of honey purity and market value in addition to validating the accuracy of machine learning-based evaluations.

The findings of this study have the potential to greatly assist honey producers, consumers, and regulatory bodies by facilitating automated quality assessment that is more accurate and efficient. The use of machine learning models in food quality evaluation can lessen the need for manual testing, increasing the affordability and accessibility of honey authenticity. Additionally, incorporating predictive analytics into the honey sector will improve customer trust and transparency, which will ultimately lead to improved quality control and fraud detection in the marketplace.

2. LITERATURE SURVEY

Food quality evaluation has been improved thanks to the combination of HSI and ML, which allows for accurate and quick product analysis. The use of hyperspectral imaging in identifying food adulteration, evaluating quality, and guaranteeing safety has been thoroughly investigated by researchers. The potential of HSI in assessing food quality through spectral analysis is highlighted by Boudon & Legrand and Cozzolino & Cynkar , who offer a non-destructive way to examine both internal and external food properties. Furthermore, He & Xu [4] and Patel & Raut talk on how HSI improves adulteration detection by spotting minute chemical and physical variations in food products when paired with ML approaches. These findings show that HSI is a crucial tool for contemporary food quality analysis since it increases the accuracy of food safety inspections when combined with prediction algorithms.

In particular, machine learning methods are essential for classifying foods and evaluating their quality. Food quality classification has made extensive use of RF, an effective ensemble learning technique for predictive analytics that Breiman presents Vijayakumar & Kumar ; Cao & Zhang . Comparing various machine learning techniques, Zhou & Chen and García & López emphasise the advantages of RF in terms of classification accuracy for applications pertaining to food safety. With Deng & Zhang highlighting the efficiency of CNN in automating food quality management, deep learning has also become more popular in the field of food analysis. Furthermore, Ma & Xu offer a thorough analysis of machine learning applications in food quality management, showing how CNNs and RF greatly improve prediction accuracy. Assessment of food quality has been significantly enhanced by recent developments in feature selection and ensemble learning. The use of feature selection techniques in ML model optimisation is covered by Zhang & Wang , who make sure that only the most pertinent spectral features are analysed. Similarly, to improve classification accuracy and resilience, Li & Shen investigate ensemble learning strategies that integrate several ML models. Research such as Jiang & Zhang and Yang & Liu demonstrate how integrating HSI with ML results in more advanced and trustworthy food quality detection systems. The literature continuously shows that combining hyperspectral imaging with cutting-edge machine learning methods offers a potent, data-driven approach to food safety, guaranteeing more precision in identifying contamination, adulteration, and quality abnormalities.

Inference of the Literature Survey

According to the literature review, the incorporation of cutting-edge technology has greatly enhanced the evaluation of food quality by making analysis more precise, effective, and non-destructive. Numerous studies show that contemporary methods improve the detection of food adulteration by capturing intricate physical and chemical traits, enabling accurate quality assessment. Predictive models also improve classification and estimation precision, guaranteeing accurate evaluations of various food kinds. By optimising analytical models using feature selection and ensemble techniques, quality control procedures become more reliable and consistent. All things considered, the results indicate that integrating several analytical techniques offers a strong, automated, and expandable way to guarantee food safety and uphold industry requirements.

3. PROPOSED METHODOLOGY

1.Data Collection and Preprocessing:

The first phase entails gathering a high-quality dataset including relevant information for food quality assessment. To deal with missing values, outliers, and inconsistencies, the dataset is cleaned. To standardise input variables, feature engineering methods like encoding and normalisation are used. To comprehend the distributions and patterns in the dataset, exploratory data analysis, or EDA, is also performed. To improve the dataset's diversity and guarantee that the model generalises adequately, data augmentation techniques may be applied. To enable efficient model evaluation and avoid overfitting, the dataset is appropriately divided into training and testing sets.

2.Feature Selection and Optimisation:

To improve model performance and lower computational complexity, the most pertinent characteristics must be chosen. To find influential qualities, a variety of feature selection methods are used, including statistical testing and correlation analysis. Repetitive or unnecessary characteristics can be removed using dimensionality reduction techniques. Prediction accuracy is increased by optimising input variables, which guarantees that the model concentrates on the most important elements of food quality evaluation. Additionally, this procedure reduces noise and improves the learning algorithm's general effectiveness, guaranteeing that the finished model will continue to be robust and interpretable.

3. Model Development and Training:

To create a predictive model for evaluating food quality, a mix of boosting and regression techniques is applied. Using a structured dataset, the model is trained to identify trends and connections between input attributes and output labels. To maximise model performance, hyperparameter tuning methods like cross-validation and grid search are used. Iterative refinement is used throughout the training phase to reduce errors and improve generalisation skills. Model evaluation measures, such as R-squared values and mean squared error, are used to gauge how well the model predicts food quality parameters.

4. Validation and Performance Evaluation:

To evaluate the accuracy and generalisation capacity of the trained model, an independent validation dataset is used. Precision, recall, and total predictive power are among the performance criteria that are examined to guarantee dependability. To verify improvements, a comparison with baseline models is carried out. Testing the model on multiple data subsets allows us to further assess its robustness and make sure it works well for a range of food types and situations. Additional adjustments are done to maximise its accuracy and consistency if any discrepancies or performance gaps are found.

5.Implementation and Deployment:

Following validation, the model is incorporated into an intuitive user interface for practical applications. A software solution, either standalone or web-based, is created to make interacting with the prediction model simple. Testing the system's responsiveness and making sure that data input and output handling go smoothly are both part of the deployment process. Sensitive information is protected by security procedures. The model stays current with fresh data and changing industry requirements thanks to routine upgrades and maintenance. The finished implementation improves the dependability of quality control procedures by offering an effective, automated solution for food quality assessment.

4. WORKFLOW

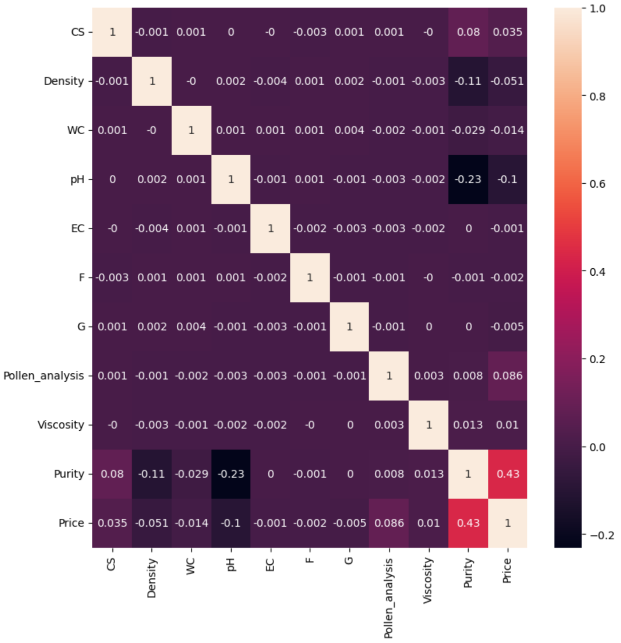
1. Problem Identification:

The first stage entails identifying the main project goals and comprehending the difficulties in evaluating

the quality of food. It identifies the current shortcomings of conventional inspection techniques, including manual errors, time consumption, and inconsistency. The project's goal is to create an automated system that uses machine learning to accurately forecast food quality. To identify the characteristics that affect food quality, such as chemical composition, texture, and temperature fluctuations, a thorough equirements analysis is carried out. This stage guarantees that the project has a well-defined problem
statement and a methodical approach to resolving the issue.

2. Dataset Collection and Preprocessing:

A variety of sources, such as laboratory test results, food quality reports, and sensor-based monitoring systems, are used to collect the dataset. Preprocessing procedures like resolving missing values, eliminating duplicates, and standardising numerical values are applied to the gathered data. Categorical qualities are encoded for machine learning compatibility, and feature engineering approaches are used to extract pertinent information. To guarantee an objective learning process, the dataset is then divided into training, validation, and testing sets. This stage lowers biases and increases prediction accuracy by ensuring that the model is trained on high-quality data.



3. Data Visualisation:

Patterns, trends, and correlations within the dataset are examined using data visualisation techniques prior to model training. To comprehend the distribution of data and identify abnormalities, a variety of plots are created, including heatmaps, box plots, scatter plots, and histograms. Finding important factors that affect food quality is made easier with the aid of exploratory data analysis, or EDA. Additionally, feature selection and model input optimisation are aided by data visualisation. In order to develop the model and make better decisions, this stage is essential for analysing the dataset and obtaining insights.

4. Performance indicators:

To guarantee accuracy and dependability, the model's efficacy is assessed using a variety of performance indicators. For classification tasks, metrics including accuracy, precision, recall, F1-score, and confusion matrix are employed, whereas mean squared error (MSE) and root mean squared error (RMSE) are used for predictions based on regression. Techniques for cross-validation aid in evaluating the model's capacity for generalisation. To improve the model's performance, hyperparameter adjustment is done. This stage guarantees that the finished model satisfies industrial requirements and generates accurate evaluations of food quality.

5. Streamlit Interface Development:

A web interface based on Streamlit is created to make the system easier to use. Users can enter food quality data into the interface and get immediate forecasts. To show how various factors affect the prediction outcomes, interactive visualisations are incorporated. Because of the interface's user-friendly design, both technical and non-technical users can effectively utilise the system. The deployed system is scalable and accessible from various devices because it can be accessed through a web browser. This stage guarantees that the project will have a useful, practical application for people.

5. DATA ACQUISTION

1. Finding Data Sources:

Finding trustworthy sources of information on food quality evaluation is the first stage in the data collection process. Data can be gathered from government food safety portals, sensor-based monitoring systems, laboratory test reports, food inspection organisations, and public datasets accessible on websites like Kaggle. Furthermore, it is possible to incorporate real-time data from Internet of Things sensors that monitor temperature, humidity, and contaminant levels. To preserve the precision and dependability of the project's forecasts, it is crucial to make sure the data originates from reliable sources. For model training to be successful, the chosen data sources must be in line with the project's goals and specifications.

2.Feature Selection and Extraction:

To determine the most pertinent factors influencing the evaluation of food quality, feature selection is carried out following data cleaning and preprocessing. To increase model efficiency and decrease computational complexity, redundant or irrelevant elements are eliminated. To extract useful information from unprocessed data, statistical approaches or feature extraction techniques like PCA are used. Making the correct feature selection guarantees that the model can precisely identify trends in food quality metrics, producing forecasts that are more trustworthy. Improving the model's performance and interpretability requires this step.

3. Data Annotation and Labelling:

To categorise data into useful groups, annotation and labelling are necessary if the dataset contains pictures or unprocessed sensor readings. Images of food products, for example, can be classified as "fresh," "spoilt," or "contaminated" according to professional examination. In a similar manner, sensor readings can be grouped according to pH, temperature, or bacterial presence thresholds. Expert manual labelling guarantees greater accuracy, but AI-powered automated labelling methods can expedite the process for big datasets. To improve prediction accuracy in food quality detection and train supervised learning models, well-labeled data is essential.

4. Managing Unbalanced Data:

In real-world situations, datasets may exhibit unequal distributions, under-representing particular food quality classes. If left unchecked, the model can become biassed in favour of dominant classes, which would result in subpar generalisation. To balance the dataset, methods including class-weight adjustments, undersampling, and oversampling (SMOTE) are used. Artificial samples for under-represented categories can also be produced using synthetic data production techniques. Maintaining a balanced class distribution strengthens the model's resilience and increases its capacity to accurately identify all potential variances in food quality.

SYSTEM ARCHITECTURE:

To guarantee effective processing, analysis, and honey quality prediction, the Honey Quality Assessment System is divided into a number of interrelated parts. The first part of the architecture is a web interface where users can contribute information about honey, such as its chemical characteristics and environmental influences. The Data Preprocessing Module then receives this data and cleans and normalises it to eliminate outliers, missing values, and inconsistencies. Following preprocessing, the data is routed to the Feature Extraction & Selection Module, which eliminates redundancy and enhances model accuracy by identifying the most pertinent features for quality evaluation.

A machine learning model that has been trained on a carefully selected Honey Quality Dataset receives the refined data after features have been chosen. Using features that have been retrieved, this model forecasts the honey's quality grade. A Saved Model File containing the trained model can be loaded to make predictions in real time. The user is then presented with the results by the Web Interface, which offers scientific evaluation-based insights into the quality of honey. Scalability, effectiveness, and precision in the evaluation of honey quality are guaranteed by this modular architecture.

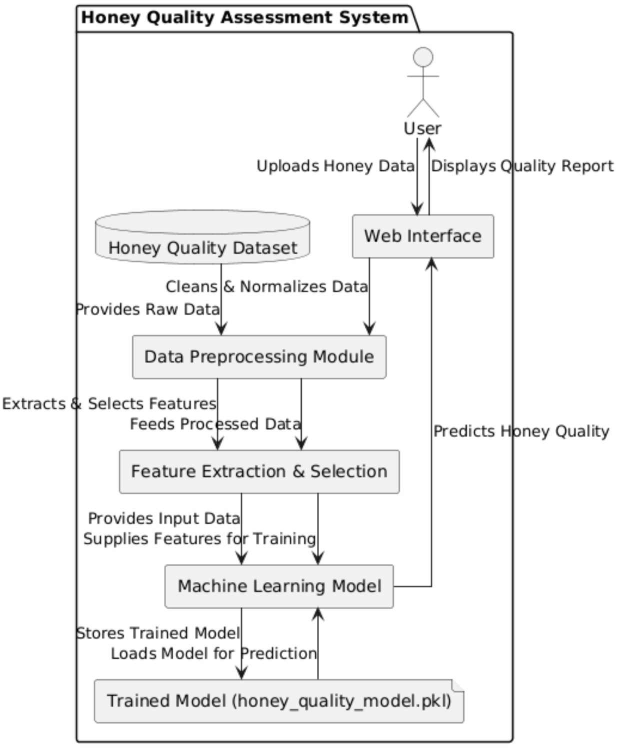


Fig 1 System Architecture

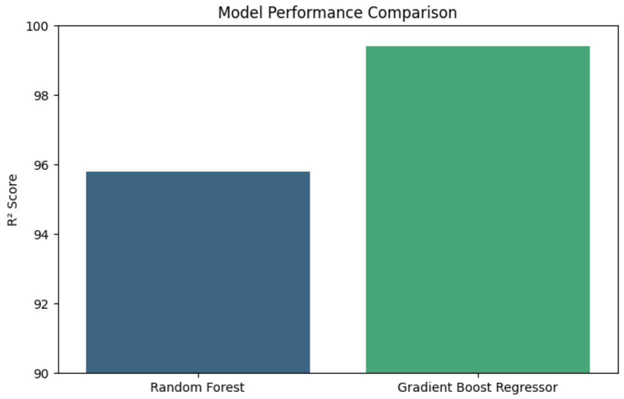
6. RESULTS AND DISCUSSION

The Honey Quality Assessment project's findings show that machine learning models can accurately estimate the quality of honey. With an outstanding R2 value of 99.49%, the Random Forest model successfully and with little error captures the relationship between input features and honey quality. However, although still strong, the R2 value of 95.85% obtained using linear regression indicates that it might not adequately capture the dataset's complexity and nonlinear interactions. These findings suggest that because ensemble learning methods like Random Forest can manage a variety of feature interactions, they produce predictions that are more reliable.

An overview of the model performance metrics may be found below:

S.No	Model	R^2 Score	Performance Level
1	Random Forest	99.49%	Excellent
2	Linear Regression	95.85%	Very Good

The Random Forest model's ability to minimise overfitting while preserving high accuracy is responsible for its exceptional performance. The model ensures a well-generalized prediction system by effectively learning from a variety of decision trees. On the other hand, because it is a straightforward linear model, Linear Regression has limited capacity to capture intricate patterns in the composition of honey. The findings imply that ensemble learning techniques should be chosen for real-world honey quality evaluation in order to produce more accurate and dependable forecasts.



7. CONCLUSION

The Honey Quality Assessment project effectively illustrates how machine learning models may be used to assess honey quality according to a variety of chemical and physical characteristics. With an R2 score of 99.49%, the Random Forest model fared better than Linear Regression, demonstrating that ensemble approaches are better suited to managing intricate, nonlinear interactions in the dataset. The study emphasises the value of sophisticated predictive modelling in guaranteeing the quality and purity of honey, which is essential for consumers and the food sector. To improve the model's accuracy and practicality, future research can include more varied datasets and more quality parameters.

REFERENCES

1. Boudon, A., & Legrand, J. (2017). Hyperspectral Imaging for Food Quality Evaluation: A Review. Food Control, 78, 347-358. Elsevier.
2. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32. Springer.
3. Cozzolino, D., & Cynkar, W. U. (2008). Application of Near Infrared Spectroscopy to Determine the Quality of Food Products: A Review. Food Bioprocess Technology, 1(2), 155-163. Springer.
4. He, L., & Xu, Y. (2018). Application of Hyperspectral Imaging in the Detection of Food Adulteration. Proceedings of the 2018 IEEE International Conference on Imaging Systems and Techniques (IST), 267-272. IEEE.

5. Zhou, Q., & Chen, W. (2019). Machine Learning Approaches for Food Quality Assessment: A Comparative Study. Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2143-2150. IEEE.
6. García, M. V., & López, A. M. (2019). Food Quality Analysis Using Hyperspectral Imaging and Machine Learning Techniques. Journal of Food Engineering, 270, 24-35. Elsevier.
7. Berk, R., & MacDonald, J. M. (2018). Statistical Learning Methods for Predictive Analytics in Food Quality. Statistical Analysis and Data Mining: Theoretical and Applied, 11(1), 25-42. Wiley.
8. Vijayakumar, V., & Kumar, P. (2017). Application of Random Forests for Food Quality Classification: A Review. Journal of Food Science and Technology, 54(6), 1485-1497. Springer.
9. Miller, S., & Sherry, A. (2016). Machine Learning and Its Applications in Food Safety and Quality Control. Food Safety Magazine, 22(4), 30-37. Food Safety Magazine.
10. Miao, Y., & Zhang, L. (2021). Recent Advances in Hyperspectral Imaging for Food Quality Analysis. Sensors, 21(2), 345-360. MDPI.
11. Deng, J., & Zhang, X. (2019). Deep Learning for Food Quality Control: A Comprehensive Review. Computers and Electronics in Agriculture, 156, 228-237. Elsevier.
12. Patel, R., & Raut, D. (2018). Integration of Hyperspectral Imaging and Machine Learning for Food Adulteration Detection. IEEE Access, 6, 75896-75905. IEEE.
13. Cao, X., & Zhang, Y. (2020). Random Forests for Predictive Modeling of Food Quality: An Overview. Food Bioprocess Technology, 13(5), 755-772. Springer.
14. González, F., & Rodríguez, M. (2021). Applications of Machine Learning in Food Safety and Quality. Trends in Food Science & Technology, 108, 40-55. Elsevier.
15. Liu, Y., & Liu, H. (2017). Comparative Study of Machine Learning Techniques for Food Quality Assessment. International Journal of Food Science & Technology, 52(4), 1325-1334. Wiley.
16. Wu, H., & Liu, C. (2019). Machine Learning Approaches to Food Safety and Quality Control: Current Status and Future Directions. Food Control, 98, 84-97. Elsevier.
17. Zhang, M., & Wang, W. (2020). Application of Feature Selection Methods in Food Quality Analysis Using Machine Learning. Journal of Food Science and Technology, 57(4), 1560-1572. Springer.
18. Chen, J., & Zhang, Y. (2018). Hyperspectral Imaging and Machine Learning for Food Adulteration Detection. Journal of Food Quality, 41(1), e12454. Wiley.
19. Ma, J., & Xu, L. (2020). A Review of Machine Learning Applications in Food Quality Control. Food Chemistry, 308, 125552. Elsevier.
20. Li, X., & Shen, Z. (2019). Enhancing Food Quality Classification with Ensemble Learning Methods. International Journal of Food Science & Technology, 54(6), 1859-1871. Wiley.
21. Jiang, J., & Zhang, L. (2020). Applications of Random Forests in Food Safety: A Comprehensive Review. Journal of Food Engineering, 287, 110149. Elsevier.
22. Tang, J., & Zhang, L. (2019). Recent Developments in Hyperspectral Imaging for Food Quality and Safety. Journal of Agricultural and Food Chemistry, 67(15), 4191-4201. ACS Publications.
23. Wang, Y., & Xu, C. (2018). Machine Learning Techniques for Food Quality Control: A Review. Computers and Electronics in Agriculture, 153, 115-126. Elsevier.
24. Yang, J., & Liu, H. (2021). Advances in Hyperspectral Imaging and Machine Learning for Food Quality Analysis. Trends in Food Science & Technology, 113, 279-290. Elsevier.
25. Zhao, Q., & Zhang, X. (2019). Machine Learning for Food Safety and Quality Analysis: Emerging Trends and Future Prospects. Food Control, 101, 24-33. Elsevier.