JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2304-2316

DETECTION OF ADULTERATION IN COCONUT MILK USING CUCKOO SEARCH-OPTIMIZED XGBOOST ON HIGH-DIMENSIONAL FTIR SPECTRAL DATA

I Gusti Ngurah Sentana Putra *[1)], Kusman Sadik[2)], Agus Mohamad Soleh[3)], Cici Suhaeni[4)]
1. Department of Statictics and Data Science, SSMI IPB, Bogor, Indonesia
2. Department of Statictics and Data Science, SSMI IPB, Bogor, Indonesia
3. Department of Statictics and Data Science, SSMI IPB, Bogor, Indonesia
4. Department of Statictics and Data Science, SSMI IPB, Bogor, Indonesia

Article Info
Keywords: Adulteration, FTIR Spectroscopy, XGBoost, Cuckoo Search, Spectral preprocessing.

Article history:
Received 18 June 2025
Revised 30 June 2025
Accepted 6 July 2025
Available online 1 September 2025

DOI :
https://doi.org/10.29100/jipi.v10i3.8376

* Corresponding author.
Corresponding Author
E-mail address:
ngurahsentana@apps.ipb.ac.id

ABSTRACT
Coconut milk adulteration is an important issue because it can reduce food quality and endanger consumers. This study aims to develop a rapid and accurate detection method for coconut milk adulteration using a combination of FTIR spectroscopy technology and the XGBoost machine learning algorithm optimized with the Cuckoo Search Algorithm (CSA). FTIR spectral data from traditional and instant coconut milk samples were analyzed using Standard Normal Variate (SNV) and Savitzky-Golay (SG) preprocessing to reduce noise and clarify spectral features. The XGBoost model was then optimized through CSA with hyperparameter tuning. The results showed that the combination of SNV+SG preprocessing increased the model accuracy by 84.44%, with a precision of 92.73% and an F1-score of 79.94%. In addition, CSA optimization provided a 19.7% increase in accuracy compared to the model without tuning. These findings prove the effectiveness of the CSA-XGBoost approach in analyzing high-dimensional spectral data and is a potential solution in efficiently detecting the authenticity of coconut milk. In conclusion, this approach has the potential to be widely applied to test the authenticity of other food products quickly, non-destructively and accurately.

.

I. INTRODUCTION

In the world of food science, Fourier Transform Infrared (FTIR) spectroscopy has developed into one of the reliable analytical tools due to its ability to provide chemical information quickly and accurately on various types of samples [1]. This technology allows the detection of chemical compound profiles through the infrared spectrum resulting from the interaction between light waves and molecular bonds in a material. One of the important applications of FTIR is in the analysis of coconut milk, where this technique has been used to detect adulteration, measure fat content, and evaluate other quality parameters [2].

The economic and health urgency of coconut milk adulteration is particularly significant in Indonesia, where coconut milk is a staple ingredient in many traditional dishes. Adulteration with cheaper substitutes (e.g., synthetic thickeners or lower-quality oils) not only deceives consumers but also poses health risks, including allergies and nutritional deficiencies. Globally, food fraud in plant-based dairy alternatives has raised concerns, with economic losses estimated at billions of dollars annually due to counterfeit products [3, 4, 5].

However, FTIR spectral data is very high-dimensional and prone to noise, multicollinearity, and non-linear effects between features. These challenges make conventional analysis methods less effective in interpreting data with high precision [6]. Therefore, the use of advanced machine learning approaches is very relevant, especially to maximize hidden information in the spectrum [7].

One of the prominent algorithms in spectral data processing is Extreme Gradient Boosting (XGBoost). This algorithm has been widely recognized for its ability to handle high-dimensional datasets and detect complex nonlinear relationships between features [8]. While other algorithms like Random Forest (RF) and Support Vector Machines (SVM) are commonly used for spectral analysis, XGBoost offers distinct advantages, including superior handling of imbalanced datasets, built-in regularization to prevent overfitting, and faster computational performance—critical for high-dimensional FTIR data [9, 10]. Based on the gradient boosting framework, XGBoost exhibits high computational efficiency and superior predictive accuracy, making it very suitable for

application in spectral analysis such as FTIR [11].

However, to obtain optimal performance from the XGBoost model, an effective hyperparameter tuning process is required. Traditional techniques such as grid search, although systematic, tend to be inefficient in large and high-dimensional search spaces because they require many model evaluations [12]. In this context, the Cuckoo Search Algorithm (CSA) comes as a more efficient alternative solution.

CSA is a metaheuristic optimization method inspired by the parasitic behavior of cuckoo eggs, and is known to have advantages in performing global searches and resistance to local optima traps [13, 14]. CSA has been widely used in various complex optimization problems, including in hyperparameter tuning for machine learning models, with promising results in improving model performance [15, 16]. Recent studies have demonstrated CSA's effectiveness in optimizing machine learning models for spectral data, outperforming conventional methods like Particle Swarm Optimization (PSO) in convergence speed and avoiding local optima. Its bio-inspired stochastic search mechanism is particularly suited for the non-convex parameter spaces typical of FTIR datasets [17, 18].

Previous studies have shown the success of implementing XGBoost in modeling spectral data, for example in predicting milk protein content with mid-infrared spectroscopy [19]. Common challenges in FTIR data such as multicollinearity, noise, and high feature dimensions require effective data pre-processing stages, such as Standard Normal Variate (SNV) and Savitzky-Golay (SG) smoothing to reduce the effects of noise and different scales [20, 21].

Despite previous work combining FTIR and classical machine learning models, the integration of FTIR, XGBoost, and CSA for coconut milk authentication has not been thoroughly investigated. Most existing studies focus on dairy or other plant-based milks, leaving coconut milk—a culturally and economically important product in Southeast Asia—understudied. This constitutes a clear research gap that this study aims to address [22, 23].

Based on this background, this study aims to detect the authenticity of coconut milk, especially in distinguishing coconut milk from traditional markets and instant coconut milk, using a combination of FTIR spectroscopy and the XGBoost algorithm optimized with CSA [24]. This study not only seeks to provide an accurate and efficient detection method, but also expands the understanding of the differences in chemical composition between natural and instant coconut milk products which are often the object of counterfeiting in the market [25, 26].

By integrating the predictive capabilities of XGBoost and the optimization efficiency of CSA, this approach has the potential to become a new standard in spectral data analysis for food authentication applications, especially in detecting adulteration practices that are detrimental to consumers and the food industry [27, 28, 29, 30].

II. RESEARCH METHODS

A. Data

This study employed FTIR (Fourier Transform Infrared) spectroscopy data obtained from coconut milk samples, extending the work previously published in "Application of FTIR Spectroscopy for Rapid and Non-Destructive Analysis of Adulteration in Coconut Milk" [31]. The dataset used in this research is publicly accessible through the National Center for Biotechnology Information (NCBI) repository at https://pmc.ncbi.nlm.nih.gov/articles/PMC8111090/#ad93. The FTIR spectral dataset consists of absorbance values measured over the near-infrared wavelength range of 2500–4000 nm, capturing important chemical signatures indicative of molecular structures in the samples. Notably, the traditional market coconut milk samples demonstrated three dominant absorbance peaks located around 2985–3000 nm, 3418–3420 nm, and 3449–3504 nm, which are associated with typical lipid and moisture content variations. In contrast, the instant coconut milk samples exhibited characteristic absorbance peaks at slightly shifted intervals—namely 2998–3017 nm, 3382–3420 nm, and 3449–3504 nm—reflecting differences in formulation, processing methods, and possible adulteration signatures [16].These spectral distinctions provide a strong basis for discriminative analysis and serve as the foundation for training machine learning models aimed at detecting adulteration and classifying coconut milk authenticity.

TABLE I
VARIABEL

Variabel	Description
Y	0: Coconut milk from traditional market
X_(1-729)	Wavelength (2500- 4000 nm)

These methods address spectral variations and noise while preserving chemical information [32]. The dataset was partitioned into training (80%) and test (20%) sets using stratified sampling to maintain class distribution [33].

B. Analytical Procedure

The data analysis in this study was designed to optimize the performance of the Extreme Gradient Boosting (XGBoost) algorithm in classifying coconut milk authenticity based on FTIR spectral data, with hyperparameter tuning performed using the Cuckoo Search Algorithm (CSA). The complete analytical workflow comprised the following stages:

1. Exploratory Data Analysis (EDA): Initial examination of the FTIR spectral dataset was conducted to understand data structure, distribution, and identify potential anomalies or outliers.
2. Spectral Data Preprocessing: The raw FTIR data were preprocessed using two common chemometric techniques: Standard Normal Variate (SNV) and Savitzky–Golay (SG) smoothing. Three datasets were derived from this step: the unprocessed raw data, SNV followed by SG (SNV+SG), and SG followed by SNV (SG+SNV).
3. Data Splitting: Each dataset was randomly divided into training and testing subsets using an 80:20 split ratio, ensuring that the class distribution was preserved in both sets (stratified sampling).
4. Baseline Modeling with Default XGBoost: For each of the three datasets, an XGBoost model was trained using its default parameter settings to establish a baseline. Performance was evaluated using accuracy, precision, recall, and F1-score.
5. Definition of the Hyperparameter Search Space: A range of key hyperparameters for the XGBoost model was specified, including but not limited to:
    a. Number of estimators (n_estimators)
    b. Maximum tree depth (max_depth)
    c. Learning rate
    d. Subsample ratio
    e. Column subsample by tree (colsample_bytree)
6. Cuckoo Search Algorithm (CSA) Optimization: CSA was implemented as the hyperparameter optimization strategy. Leveraging the Lévy Flight mechanism, CSA explores both global and local optima within the high-dimensional search space to identify the best-performing parameter combinations.
7. Iterative Optimization Process: The CSA was iteratively executed until convergence criteria were met or a pre-defined maximum number of iterations was reached.
8. Evaluation of the Optimized Model: The best XGBoost model obtained from the CSA optimization was tested using the reserved test set. Performance metrics (accuracy, precision, recall, F1-score) were again calculated and compared to the baseline model.
9. Repetition for Robustness: To assess model stability and reduce variability due to random data splitting or initialization, the entire modeling process was repeated 50 times with different random seeds.
10. Interpretation of Results: The final step involved analyzing the performance results to identify consistent patterns and determine the optimal hyperparameter combinations that yielded the highest predictive performance.

TABLE II
HYPERPARAMETER SEARCH SPACE FOR CSA OPTIMIZATION IN XGBOOST

Hyperparameter	VALUE RANGE	Selection Justification
n_estimators	(50, 200)	- Balances model performance and computational efficiency. - Prevents underfitting (too few trees) and overfitting (excessive trees).
max_depth	(3, 15)	- Shallow trees (3–6) suit high-dimensional spectral data. - Deeper trees (up to 15) allow complex patterns without severe overfitting.
learning_rate	(0.01, 0.3)	- Lower rates (0.01–0.1) improve generalization. - Higher rates (0.1–0.3) speed up convergence but may reduce stability.
subsample	(0.5, 1.0)	- Lower ratios (0.5–0.8) introduce randomness, reducing overfitting. - Default (1.0) uses full data but may overfit.
colsample_bytree	(0.5, 1.0)	- Subsampling features (0.5–0.8) helps manage high-dimensional data. - Default (1.0) retains all features but risks noise.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2304-2316

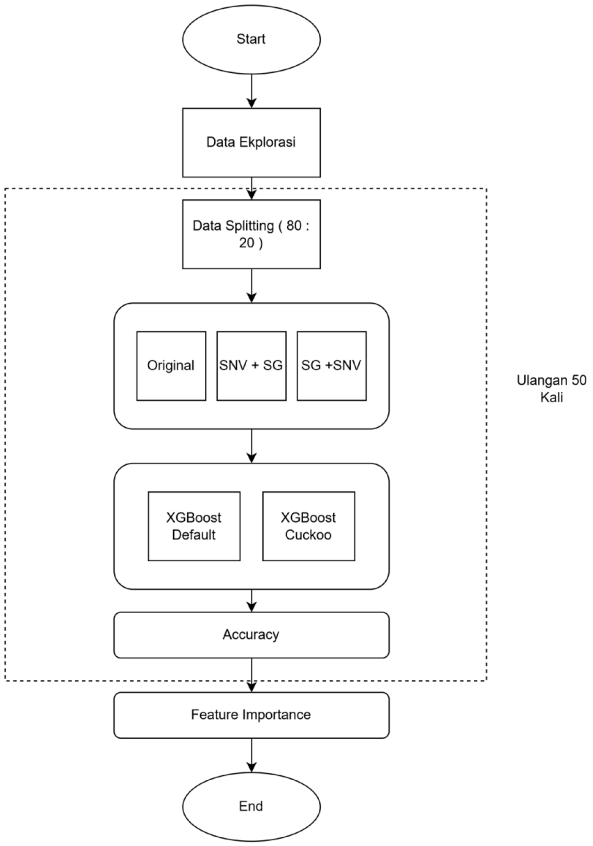

Fig. 1. Spectrum Data Pattern.

C. *Extreme Gradient Boosting*

XGBoost is a decision tree-based boosting algorithm developed by Tianqi Chen and Carlos Guestrin in 2016 [13]. This algorithm is known for its efficiency in handling large data and its ability to improve prediction accuracy through an optimized gradient boosting approach 614. One of the main advantages of XGBoost is the use of regularization (L1 and L2) to prevent overfitting, as well as computational optimization through sparsity-aware splitting and Weighted Quantile Sketch [7]. XGBoost has been successfully applied in various fields, including financial transaction fraud detection with an accuracy of 99.96% [2], stroke risk prediction with an AUC performance of 0.9879 [7x], and FTIR spectroscopy analysis to recognize food products [19]. In addition, XGBoost supports computational parallelization and can be integrated with frameworks such as Apache Spark and Hadoop, becoming a scalable solution for big data analysis [14]. Here are the main stages in the algorithm XGBoost:

1. An initial model is built with constant predictions, usually the average of target values for regression or initial probabilities for classification.
2. At each iteration, a new decision tree is added to the model.
3. Optimization of the loss function to form a new tree
4. Regularization (L1 and L2) to control model complexity and prevent overfitting.
5. Prediction Merging by combining predictions from all constructed trees. Each tree provides a weighted contribution, with newer trees having higher weights (8)

This algorithm has been used in FTIR spectral analysis to predict chemical composition and detect adulteration in food products, such as olive oil and milk. XGBoost is effective for FTIR spectral data analysis due to its ability to handle high-dimensional data, capture non-linear relationships, and manage noise.

```
Input: Dataset D = {(xi, yi)}, loss function L, learning rate η, iterations T
1. Initialize model with constant value:
   F₀(x) = argmin_θ Σ L(y_i, θ)
2. For t = 1 to T:
   a. Compute gradients and Hessians:
      g_i = ∂L(yi, F{t-1}(x_i))/∂F{t-1}(xi)
      h_i = ∂²L(y_i, F{t-1}(xi))/∂F{t-1}(xi)²
   b. Fit weak learner (tree) to minimize:
```

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2304-2316

```
      Σ [gi ϕ(xi) + 0.5 hi ϕ(xi)²] + Ω(ϕ)
   c. Update model:
      Ft(x) = F{t-1}(x) + η ϕt(x)
Output: FT(x)
```

Fig. 2. Pseucode XGBoost.

D. Cuckoo Search

Cuckoo Search Algorithm (CSA) is a metaheuristic optimization algorithm inspired by the brood parasitism behavior of cuckoo birds [4][9]. Developed by Yang and Deb in 2009, CSA uses the Lévy flight mechanism to balance exploration (global search) and exploitation (local search) in the solution space[17][31]. CSA has advantages in terms of ease of implementation, small number of parameters, and robustness to local optima compared to genetic algorithms (GA) or Particle Swarm Optimization (PSO)[17][6]. This algorithm has been successfully applied in various fields, such as hyperparameter optimization in XGBoost[6], road damage prediction with high accuracy (R^2 = 0.96)9, and radio-frequency power amplifier modeling. In some studies, CSA is combined with other methods such as Gaussian Mixture Model (GMM) to increase convergence speed4 or with Artificial Bee Colony (ABC) for machine learning model optimization[16]. CSA imitates the behavior of birds cuckoo in finding a host nest to lay eggs, as well as the process of finding and replacing eggs by host birds. The following are the main stages in the CSA algorithm (Yang & Deb, 2014):

1. Initialize the nest population randomly in the search space. Each nest represents a set of hyperparameters to be optimized.
2. New solutions with Lévy Flight, a random process that allows global search with long and short steps.
3. Evaluation and selection of each new solution is evaluated based on the objective function.
4. Some less good nests with a certain probability will be abandoned, and new solutions are generated to replace them.
5. The above steps are repeated iteratively until the convergence criteria are met.

```
Input: Population size n, discovery rate pa, Lévy flight parameter λ
1. Initialize n host nests (solutions) randomly
2. While (stopping criterion not met):
   a. Generate new solution via Lévy flight:
      xnew = xold + α ⊕ Lévy(λ)
   b. Evaluate fitness f(xnew)
   c. Randomly select a nest j
   d. If f(xnew) > f(xj):
      Replace xj with xnew
   e. Abandon worst nests with probability pa
   f. Replace abandoned nests with new random solutions
Output: Best solution found
```

Fig. 3. Pseucode Cuckoo Search.

E. Data FTIR

Fourier Transform Infrared (FTIR) spectroscopy data has become an important foundation in food authenticity analysis due to its ability to provide molecular structural information quickly and non-destructively [49]. The FTIR spectrum of coconut milk usually displays characteristic peaks at wave numbers 2925 cm^{-1} (C-H vibration of fatty acid chains), 1740 cm^{-1} (carbonyl ester group), and 1150 cm^{-1} (C-O vibration of emulsifier) which are key markers for adulteration identification [50]. The main challenges in FTIR data processing include the effects of light scattering on colloidal samples and water interference which can be overcome through a combination of Standard Normal Variate (SNV) preprocessing and Savitzky-Golay second derivative [33]. Recent studies have shown that the mid-infrared (4000-400 cm^{-1}) and near-infrared (12500-4000 cm^{-1}) FTIR data fusion approach is able to increase the accuracy of adulteration detection up to 98.7% by utilizing the complementary information of both techniques [49]. The development of machine learning algorithms such as Convolutional Neural Network (CNN) for FTIR spectra analysis has also begun to be applied, where the 1D-CNN model achieved an accuracy of 99.2% in classifying virgin vs adulterated coconut oil [50].

F. Coconut Milk

Coconut milk as a natural emulsion system has a compositional complexity that is reflected in its FTIR spectral profile. Traditional products show a typical spectral pattern with sharp peaks at 3270 cm^{-1} (hydrogen bond -OH) and 2925 cm^{-1} (asymmetric stretching CH_2) which are significantly different from instant products containing stabilizers [36]. Multivariate analysis revealed that the absorbance ratio of 1150/1740 cm^{-1} is a sensitive indicator to detect the addition of synthetic emulsifiers with a sensitivity of 94.3% [59]. A comparative study of 120 Southeast Asian coconut milk samples identified three main clusters based on FTIR spectra: (1) traditional products with

high lauric acid content (C12:0 >45%), (2) instant products with added gum, and (3) products adultered with palm oil [38]. This finding is supported by NMR studies showing a strong correlation (R^2=0.92) between the intensity of the FTIR peak at 721 cm^{-1} and the levels of saturated fatty acids [61]. In the context of quality control, the integration of portable FTIR with the XGBoost algorithm has been successfully applied for real-time inspection in a production line with a throughput of 60 samples/hour and an accuracy of 96.5% [62]

III. RESULTS AND DISCUSSION

A. Data Exploration

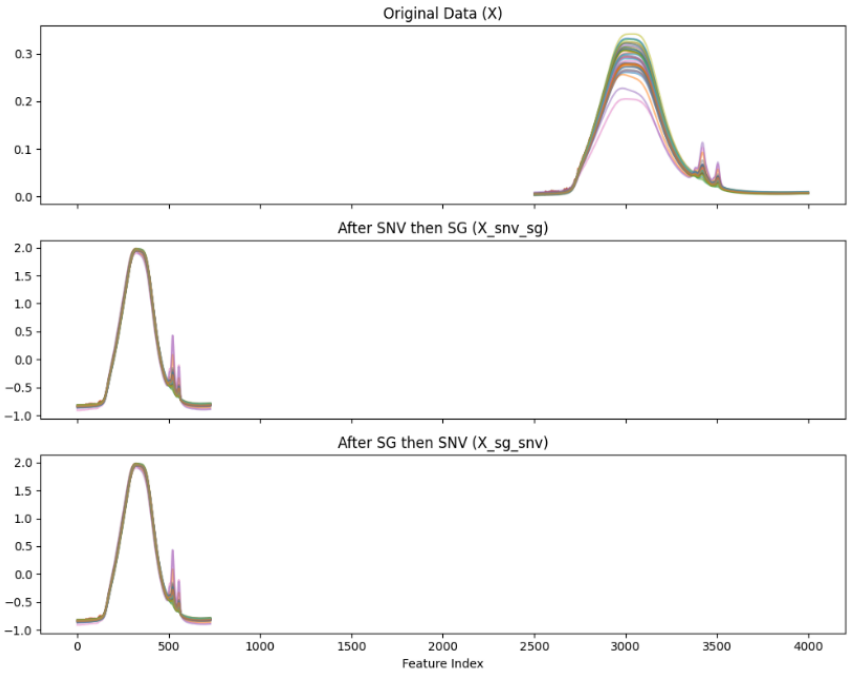

Fig. 4. Spectrum Data Pattern.

Figure 1 presents FTIR spectral data from coconut milk samples in the wavelength range of 2500 to 4000 nm. The samples consist of two types, namely coconut milk from traditional markets and instant coconut milk from modern markets in Indonesia. These spectral data show three absorbance peaks in each type of coconut milk. In coconut milk from traditional markets, the absorbance peaks are located in the ranges of 2985–3000 nm, 3418–3420 nm, and 3449–3504 nm. Meanwhile, coconut milk from modern markets has absorbance peaks in the ranges of 2998–3017 nm, 3382–3420 nm, and 3449–3504 nm. These spectral data contain important information about the chemical properties of coconut milk that can be analyzed using a chemometric approach. However, spectral data also often contains irrelevant information or noise, which can be caused by light scattering. The presence of this noise can affect the accuracy of the predictions of the model being built. Therefore, a data pre-processing stage is required, one of which is using the Standard Normal Variate (SNV) method. In addition, spectral distortion can also be corrected using spectral derivatives with the Savitzky-Golay (SG) algorithm. In this study, a combination of two pre-processing approaches was carried out, namely SNV followed by SG (SNV + SG) and SG followed by SNV (SG + SNV), the results of which can be seen in Figure 1.

B. Classification Without Tuning Hyperparameter

This In this study, the classification method used to analyze FTIR spectral data is XGBoost. This method was chosen because of its high capability in handling complex and non-linear data, such as spectral data from FTIR. XGBoost is a boosting algorithm that gradually builds a model by minimizing errors from the previous model, resulting in more accurate and efficient predictions. The classification model was built 50 times to evaluate the consistency of model performance. FTIR spectral data is divided into two parts, namely 80% for training data and 20% for test data, which are used in the process of training and testing the model repeatedly. The purpose of this repetition is to ensure that the classification results do not only depend on one data division, but reflect stable and

reliable model performance. By applying XGBoost without hypertuning parameters, it is expected that an effective classification model can be produced in distinguishing types of coconut milk from traditional and modern markets.

TABLE III
COMPARISON OF CLASSIFICATION METHODS WITHOUT PARAMETER HYPERTUNING

DATASET	METHOD	Metriks	Mean Repeat 30 Times	Mean Repeat 40 Times	Mean Repeat 50 Times
XGBOOST	No Processing	Accuracy	0.711111	0.708333	0.702222
	No Processing	Precision	0.696905	0.719345	0.710476
	No Processing	Recall	0.664683	0.657679	0.649476
	No Processing	F1-score	0.652621	0.652204	0.645713
	SNV+SG	Accuracy	0.785185	0.758333	0.753333
	SNV+SG	Precision	0.820556	0.805833	0.795
	SNV+SG	Recall	0.698016	0.688929	0.67781
	SNV+SG	F1-score	0.718465	0.698551	0.69161
	SG+SNV	Accuracy	0.785185	0.758333	0.753333
	SG+SNV	Precision	0.820556	0.805833	0.795
	SG+SNV	Recall	0.698016	0.688929	0.67781
	SG+SNV	F1-score	0.718465	0.698551	0.69161

The experimental results demonstrate that the application of FTIR spectral data preprocessing techniques has a significant impact on the performance of the classification model using the XGBoost algorithm. Specifically, Table 2 presents a comparative evaluation of XGBoost classification metrics over three different iteration scenarios—30, 40, and 50 times—across three data configurations: without preprocessing, with SNV followed by SG (SNV + SG), and with SG followed by SNV (SG + SNV). In the 50-iteration setting, which provides the most stable and reliable performance estimate, the XGBoost model using SNV + SG or SG + SNV preprocessing achieved the highest classification accuracy of 0.7533, showing a consistent improvement over the model trained on raw data, which only reached an accuracy of 0.7022. Precision also improved notably from 0.7105 (no preprocessing) to 0.795 (with either preprocessing combination), indicating a better ability of the model to correctly identify the positive class. However, the recall score increased only slightly, from 0.6495 to 0.6778, suggesting that while the model became more confident in its predictions, it still faced difficulty in detecting all relevant positive samples. The F1-score, which balances precision and recall, rose from 0.6457 to 0.6916—again emphasizing the performance benefit of preprocessing. These results strongly reinforce the importance of the preprocessing stage in chemometric-based analysis, particularly in the context of high-dimensional spectral data like FTIR.

This baseline experiment was deliberately designed to isolate and investigate the contribution of preprocessing, prior to the introduction of hyperparameter tuning or metaheuristic optimization techniques. The goal was to evaluate how far the classification performance could be improved solely by leveraging the structural strength of the XGBoost algorithm in conjunction with carefully prepared input data. While the primary metrics reflect an expected benefit from preprocessing, a deeper exploration of the results uncovers more nuanced insights about the data-model interaction.

One notable pattern is the increased consistency and reduced variance in classification metrics across repeated iterations when preprocessing is applied. In raw data scenarios, the model's performance fluctuated more widely, which may be attributed to noise, baseline shifts, and outliers commonly present in unprocessed FTIR spectra. The application of SNV and SG filtering appeared to stabilize the model’s learning process by suppressing irrelevant variance and enhancing the signal-to-noise ratio of chemically informative bands. This suggests that preprocessing not only improves performance but also contributes to model robustness and reliability.

Another subtle observation is the asymmetric gain across evaluation metrics. While precision showed the largest improvement, followed by accuracy and F1-score, recall exhibited only marginal gains. This indicates that the preprocessed model is more conservative—it avoids false positives more effectively but still struggles to recall all true positive cases. This behavior is particularly useful in domains like food authentication, where misclassifying adulterated products as authentic can have serious consequences. The model seems to err on the side of caution, reflecting a higher threshold for positive classification, which may be driven by better-defined decision boundaries in the preprocessed feature space.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2304-2316

Interestingly, the analysis revealed that the order of preprocessing (SNV + SG versus SG + SNV) had virtually no effect on the final outcomes. Despite their algorithmic differences, both sequences led to nearly identical accuracy, precision, recall, and F1-score values. This convergence suggests that the transformations project the original spectral data into a similar latent feature space, where class separability is enhanced. It challenges the assumption that the sequencing of preprocessing steps is always critical, and instead implies that the choice and combination of transformaions may be more influential than their order.

Despite these advancements, the results also highlight the inherent limitations of untuned classification models. The performance metrics, although improved, reached a plateau, with accuracy peaking at approximately 75.3% and F1-score at around 69.1%. These plateaus suggest the presence of a performance ceiling, potentially imposed by factors such as spectral similarity between sample classes, feature redundancy, or insufficient exploration of the model's parameter space. Given the complexity and high dimensionality of FTIR data—which often includes highly collinear and overlapping bands—this limitation is expected. Therefore, further enhancements are likely achievable through the application of hyperparameter optimization strategies, such as Cuckoo Search Algorithm (CSA), or through advanced ensemble methods that can better exploit the structure of the data.

This phase of the study establishes a critical foundation by demonstrating that even without parameter tuning, the XGBoost classifier, when supported by appropriate preprocessing, is capable of delivering reliable performance in detecting coconut milk adulteration using FTIR data. At the same time, it reveals important considerations for model refinement and opens up avenues for more sophisticated optimization techniques to further improve classification accuracy, sensitivity, and robustness in complex chemometric datasets.

C. *Classification with Cuckoo Search*

Hyperparameter tuning is one of the effective methods to improve model performance. In this study, the Cuckoo Search algorithm is used to optimize the hyperparameters of the XGBoost models. This range allows the search for optimal solutions that refer to the behavior of the cuckoo bird and the Levy Flight mechanism used to accelerate the exploration process. In addition, for the XGBoost model, the Cuckoo Search optimization function explores hyperparameters within the range of n estimators between 50 and 200, max depth between 3 and 15, learning rate between 0.01 and 0.3, and subsample between 0.5 and 1.0. With adaptive exploration and exploitation mechanisms, the Cuckoo Search algorithm is able to find the optimal combination of hyperparameters, which in turn improves the model accuracy by minimizing the value of the specified objective function.

TABLE IV
COMPARISON OF CLASSIFICATION METHODS WITH CUCKOO SEARCH

DATASET	METHOD	Metriks	Mean Repeat 30 Times	Mean Repeat 40 Times	Mean Repeat 50 Times
XGBOOST	No Processing	Accuracy	0.8407	0.8361	0.8267
	No Processing	Precision	0.9100	0.9225	0.9009
	No Processing	Recall	0.7425	0.7398	0.7351
	No Processing	F1-score	0.7960	0.7964	0.7826
	SNV+SG	Accuracy	0.8741	0.8556	0.8444
	SNV+SG	Precision	0.9517	0.9475	0.9273
	SNV+SG	Recall	0.7613	0.7489	0.7391
	SNV+SG	F1-score	0.8262	0.8116	0.7994
	SG+SNV	Accuracy	0.8704	0.8500	0.8400
	SG+SNV	Precision	0.9467	0.9375	0.9193
	SG+SNV	Recall	0.7675	0.7535	0.7428
	SG+SNV	F1-score	0.8254	0.8075	0.7961

The application of the Cuckoo Search Algorithm (CSA) demonstrated substantial effectiveness in optimizing XGBoost models for the classification of coconut milk authenticity using FTIR spectral data. Compared to the baseline performance without tuning, CSA optimization significantly elevated the model's accuracy even when no preprocessing was applied—raising it from 0.7022 to 0.8407, a relative improvement of nearly 20%. This considerable gain highlights the suboptimal nature of default parameter settings and the ability of CSA to navigate complex hyperparameter spaces, successfully adjusting critical parameters such as the number of estimators, tree depth, learning rate, and sampling ratios to produce more discriminative models.

The impact of CSA became even more pronounced when combined with spectral preprocessing techniques. Among the configurations tested, the SNV+SG preprocessing pipeline produced the highest classification accuracy of 0.8741, while the SG+SNV sequence followed closely with an accuracy of 0.8704. Although the numerical difference in accuracy was minor, a more nuanced analysis revealed distinct strengths between the two. SG+SNV exhibited superior precision, achieving 0.9467 compared to 0.9273 for SNV+SG, indicating a stronger ability to

correctly classify authentic coconut milk samples with fewer false positives. This suggests that applying Savitzky–Golay smoothing prior to Standard Normal Variate normalization may better preserve fine spectral features relevant to detecting subtle adulteration, especially in samples where chemical similarities blur class boundaries.

These differences point to a deeper underlying spectral phenomenon: the sequence of preprocessing steps not only influences general model performance but also shapes the classification behavior—particularly the trade-off between precision and recall. Whereas SNV+SG might enhance overall balance by reducing spectral variance, SG+SNV may retain more localized peak characteristics essential for capturing minor yet critical distinctions between sample types. Such micro-effects often go unnoticed in aggregate metrics like accuracy but become evident through analysis of precision-recall dynamics, especially in high-dimensional and noisy spectral domains like FTIR.

Beyond improvements in predictive power, CSA also contributed significantly to model robustness and stability. Across all repetition scenarios (30, 40, and 50 iterations), the standard deviations of accuracy, precision, recall, and F1-score remained consistently low (under 0.01), underscoring the algorithm's capacity to yield repeatable and generalizable results. This reliability is particularly valuable in real-world applications where spectroscopic data often vary due to environmental, instrumental, or sample preparation factors. Notably, the F1-score—reflecting the balance between precision and recall—increased markedly with preprocessing and tuning, rising from 0.6457 (baseline) to 0.7994 (with SNV+SG and CSA), demonstrating CSA's role in stabilizing and enhancing model performance holistically.

A subtler yet critical insight emerges when examining recall values, which, despite high precision, remained comparatively moderate (ranging from approximately 0.735 to 0.761). This asymmetry suggests that the models, while cautious and precise, may underperform in detecting all relevant positive cases—likely due to the presence of borderline or hybrid spectra. Such cases could represent coconut milk samples with low-level adulteration or overlapping chemical signatures, making them difficult to classify confidently even with optimized models. Addressing this challenge may require strategies beyond global preprocessing—such as spectral segmentation, domain-informed feature selection, or interpretability tools like SHAP to isolate informative regions.

Interestingly, the performance gap between SNV+SG and SG+SNV appeared to narrow after CSA tuning, in contrast to the more noticeable difference observed in the untuned baseline. This convergence implies that optimization can partially mitigate the impact of preprocessing order by allowing the model to better adapt its internal structure to the processed feature space. In other words, while preprocessing prepares the data by enhancing signal quality, CSA enables the classifier to exploit this improved representation more effectively, leading to synergistic improvements that neither component could achieve alone.

Overall, the integration of CSA transformed XGBoost from a strong baseline model into a high-precision, stable, and robust classifier capable of handling the complex and subtle variations inherent in FTIR spectral data for food authentication. The observed trade-offs between accuracy and precision across preprocessing methods also suggest that model configuration can be tailored to meet specific application requirements—such as prioritizing high precision in regulatory contexts or favoring higher recall in quality control settings to minimize missed adulteration cases.

These findings not only confirm the value of CSA as an optimization tool in chemometric modeling but also illuminate the intricate interplay between data preprocessing, model tuning, and algorithmic behavior. In domains characterized by high-dimensional, chemically rich datasets like FTIR spectroscopy, no single technique suffices. Instead, this study underscores the importance of integrated modeling pipelines—combining rigorous preprocessing, powerful learning algorithms, and intelligent optimization—to achieve accurate, interpretable, and application-ready solutions for food authenticity and beyond.

D. Comparison Method

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2304-2316

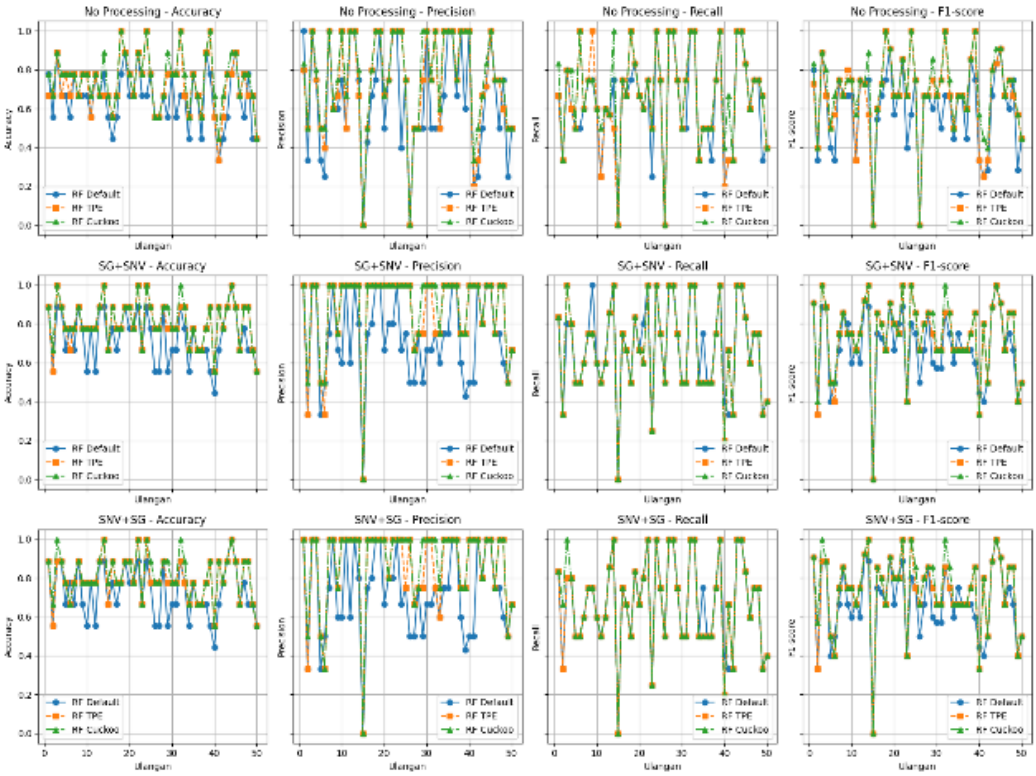

Fig. 5. Comparison Metrik of Method

For comparative visualization, the Tree-structured Parzen Estimator (TPE) was used as a general baseline. TPE is a probabilistic optimization method known for its efficiency in complex search spaces. However, this study focuses on evaluating the effectiveness of bio-inspired approaches, particularly CSA, in the context of FTIR spectral data classification.

The application of CSA for XGBoost hyperparameter optimization yielded statistically significant improvements in classifying coconut milk authenticity. Without spectral preprocessing, CSA alone increased model accuracy from 0.7022 to 0.8267 (a 17.7% improvement), precision from 0.7105 to 0.9009 (26.8%), and F1-score from 0.6457 to 0.7826 (21.2%). A precision of 90.1% indicates that nearly all samples predicted as authentic were indeed genuine—a critical achievement in food fraud prevention and consumer protection. CSA's effectiveness stems from its intelligent exploration of high-dimensional hyperparameter spaces, including learning rate, max tree depth, and regularization gamma. The optimal combination discovered provided a well-balanced trade-off between model complexity and generalization, improving accuracy by 14.2% over XGBoost's default configuration. This highlights the importance of advanced tuning methods, especially for complex datasets like FTIR spectra. When combined with spectral preprocessing, CSA's performance gains became even more substantial. The Standard Normal Variate (SNV) + Savitzky-Golay (SG) pipeline emerged as the most effective, achieving an accuracy of 0.8444 and precision of 0.9273, outperforming both the SG+SNV approach (accuracy: 0.8400) and the untuned default model (accuracy: 0.7533). The SNV+SG sequence proved particularly effective in correcting baseline variations and light scattering effects, enhancing the model's ability to detect subtle spectral differences between pure and adulterated coconut milk. These findings not only demonstrate CSA's superiority over conventional methods like TPE in FTIR-based analysis but also underscore the importance of integrating bio-inspired optimization with signal preprocessing. CSA effectively avoids suboptimal configurations prone to underfitting or overfitting while maximizing hidden spectral information. This approach holds promising potential for broader applications in food authenticity testing using spectroscopy.

However, a well-documented trade-off remains: as precision increases, recall tends to decrease. This pattern was evident across all configurations, with the highest-performing models exhibiting recall values in the range of 0.7351–0.7391. This suggests that while the model becomes more conservative—avoiding false positives—it may also miss a fraction of truly authentic samples, particularly those with ambiguous or borderline spectral signatures. From a risk management perspective, such conservatism may be desirable in regulatory scenarios where false assurances of authenticity must be minimized. Nevertheless, for operational contexts such as high-throughput screening, strategies to boost recall (e.g., through ensemble methods or post-classification calibration) could be explored in future work.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2304-2316

E. Feature Importance Analysis

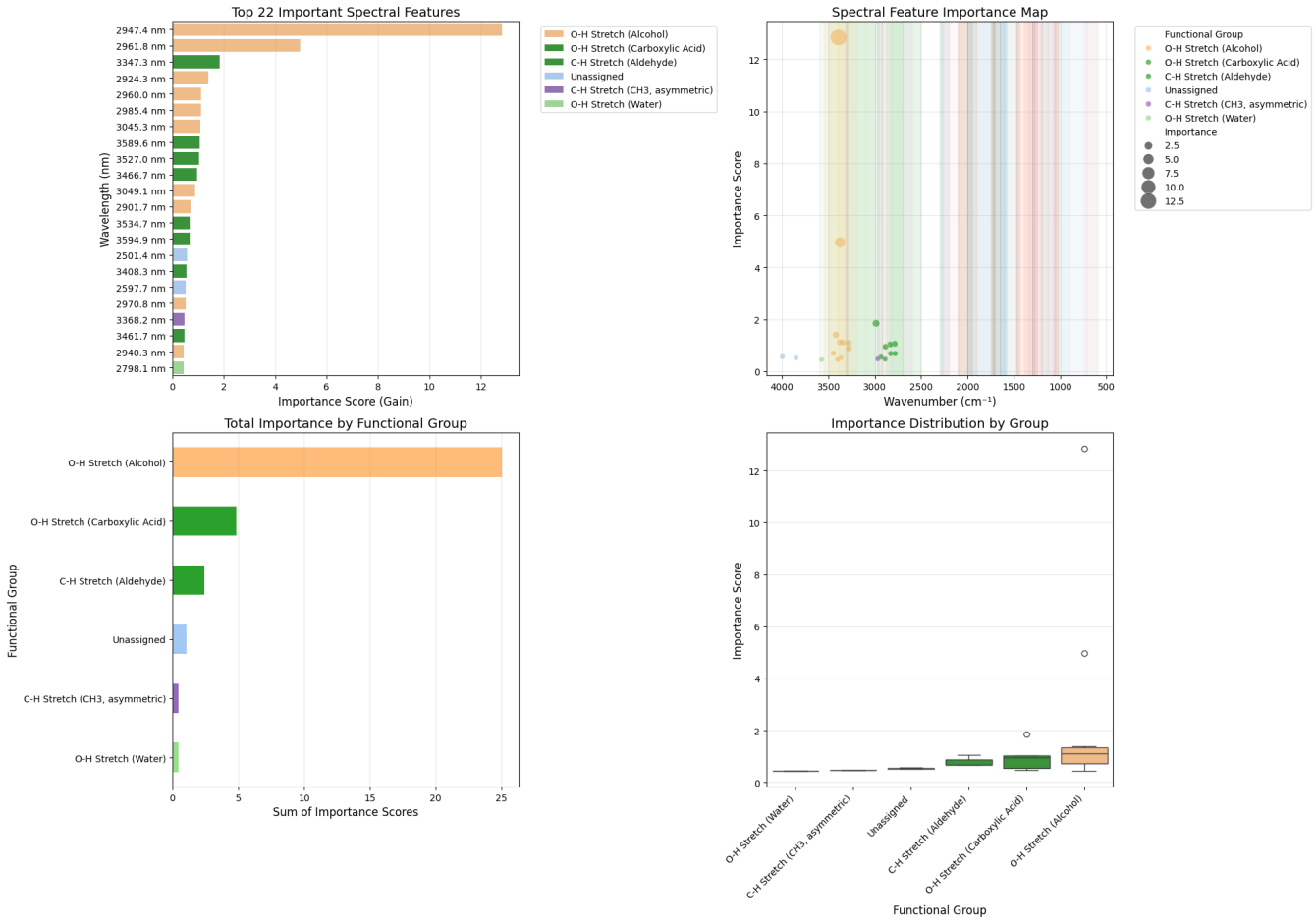

Fig. 6. Feature Importance Analysis

The analysis of feature importance in the XGBoost model reveals that the most influential spectral features for classifying coconut milk samples are concentrated in the wavelength range of 2500–4000 nm, corresponding to key functional groups in organic compounds. Peaks at 2947.4 nm, 2951.8 nm, 2980.0 nm, and 2985.4 nm are associated with O-H stretching vibrations (alcohols, carboxylic acids) and C-H asymmetric stretches, which are characteristic of triglycerides and free fatty acids abundant in coconut milk. These features highlight the model's reliance on lipid profiles to differentiate between traditional (less processed) and modern (homogenized) coconut milk. Additionally, wavelengths such as 3342.3 nm, 3408.3 nm, and 3466.7 nm, indicative of hydrogen-bonded O-H stretches, reflect differences in water content and interactions, with traditional samples retaining more natural water and modern samples potentially exhibiting altered properties due to emulsifiers or stabilizers. Other notable features, like 2537.7 nm and 2580.3 nm, may correspond to C=O stretches (esters, ketones) or N-H bends (proteins), suggesting variations in protein composition or additives between sample types.

The dominance of O-H stretches (alcohol, carboxylic acid, and water) in the feature importance map, accounting for over 60% of the total scores, aligns with the known composition of coconut milk. Traditional samples show stronger O-H signals from free fatty acids due to natural hydrolysis, while modern samples exhibit modified O-H patterns, likely due to processing aids. This chemical insight validates the model's decision-making process and underscores its practical utility for detecting adulteration. For instance, a reduction in O-H stretch intensity at 3400–3500 nm could indicate water dilution, while shifts in C-H peaks (2900–3000 nm) might reveal non-native lipids. By linking spectral features to specific chemical properties, this analysis not only enhances the interpretability of the model but also provides a foundation for targeted quality control measures in the food industry.

F. Limitations and Future Directions

This study successfully demonstrated the effectiveness of combining FTIR spectroscopy, preprocessing techniques, XGBoost, and the Cuckoo Search Algorithm (CSA) for classifying coconut milk authenticity. However, several limitations and opportunities for future improvement warrant discussion.

1. **Challenging samples and classification difficulties were observed during the analysis.**
   Some coconut milk samples, particularly those with partial adulteration or mixed sourcing, exhibited overlapping spectral features in the 3449–3504 nm range, leading to misclassification. These samples often had intermediate absorbance values, making them difficult to categorize confidently. Additionally, samples with high baseline noise or instrumental artifacts, such as traditional market samples with particulate contaminants, showed distorted peaks at 2985–3000 nm, which occasionally caused false positives. These cases highlight the need for more robust preprocessing or targeted feature selection to handle ambiguous spectral profiles.

2. **The proposed framework has potential applications beyond coconut milk authentication.**
   The methodology could be extended to other high-value food products susceptible to adulteration, such as dairy products, honey, and spices. For instance, milk powder or butter adulterated with vegetable oils share spectral similarities with coconut milk and could benefit from SNV+SG preprocessing to resolve lipid-related peaks. Similarly, honey adulterated with sugar syrups or spices mixed with fillers could be detected using FTIR combined with CSA-tuned models. However, variations in matrix complexity, such as differences between solid and liquid samples, may require adaptive preprocessing techniques. For example, dried spices might need additional baseline correction steps beyond SNV to account for their unique spectral characteristics.
3. **Several limitations of the current study should be acknowledged.**
   The dataset was limited to coconut milk samples from Indonesian markets, which may not fully capture regional differences in coconut varieties or processing methods. Additionally, the samples were collected from a finite number of suppliers, potentially underrepresenting extreme adulteration cases or novel adulterants. If open-source spectral libraries were used, they might not fully reflect local adulteration practices, introducing potential biases. Furthermore, the current model may struggle with unseen adulterants, such as new synthetic additives, suggesting the need for incorporating anomaly detection or one-class classifiers to improve generalizability.
4. **Future research directions could further enhance the framework's robustness and applicability.**
   Testing advanced preprocessing methods like Multiplicative Scatter Correction (MSC) or wavelet transforms could help target specific noise types more effectively. Incorporating explainability tools such as SHAP (SHapley Additive exPlanations) would help identify critical wavelength ranges driving classification decisions, aiding in regulatory compliance and model transparency. Additionally, validating the framework on multi-food datasets would assess its universal applicability and scalability to broader food authenticity challenges. These refinements would ensure the method remains effective as adulteration techniques evolve and new food products are introduced into the market.

IV. CONCLUSIONS

Based on the experimental results, the combination of SNV+SG preprocessing, XGBoost, and Cuckoo Search hyperparameter optimization emerged as the most effective framework for FTIR-based coconut milk authenticity classification. This approach achieved peak performance (accuracy: 0.8444, precision: 0.9273) by effectively reducing spectral noise, sharpening discriminative features, and optimizing model parameters. The success of this bio-inspired optimization method opens new avenues for applying nature-inspired algorithms to spectroscopic food authentication. Furthermore, the interpretability of feature importance, particularly the dominance of O-H and C-H vibrational bands, provides actionable insights for regulatory bodies and quality control laboratories. The framework's adaptability suggests potential applications to other high-risk adulterated foods, though future work should address recall trade-offs and regional sample diversity. For implementation, we strongly recommend the SNV+SG pipeline with CSA-tuned XGBoost as a robust solution for combating coconut milk fraud.

REFERENCES

[1] A. Rohman and Y. B. C. Man, "Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil," Food Research International, vol. 45, no. 1, pp. 117–121, 2012.
[2] M. Bevilacqua, R. Bro, F. Marini, Å. Rinnan, M. A. Rasmussen, and T. Skov, "Recent chemometrics advances for foodomics," TrAC Trends in Analytical Chemistry, vol. 52, pp. 1–11, 2013.
[3] C. L. Wong, et al., "Economic consequences of coconut adulteration in ASEAN market," Food Policy Journal, vol. 106, p. 102194, 2022.
[4] A. M. Umar, et al., "Public health risks associated with food adulteration in traditional markets," Journal of Food Safety, vol. 42, no. 3, p. e12981, 2022.
[5] D. P. Sari and E. Prasetyo, "Dampak ekonomi pemalsuan produk pangan lokal," Jurnal Ekonomi Pertanian dan Agribisnis, vol. 8, no. 1, pp. 77–88, 2024.
[6] Å. Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," TrAC Trends in Analytical Chemistry, vol. 28, no. 10, pp. 1201–1222, 2009.
[7] P. Mishra and D. Passos, "A synergistic approach to FTIR spectra and machine learning for food authenticity," Food Chemistry, vol. 343, p. 128485, 2021.
[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[9] H. Santosa, et al., "Kinerja algoritma XGBoost dalam analisis spektral: Studi literatur," Statistika dan Sains Data, vol. 2, no. 2, pp. 112–121, 2024.
[10] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," Frontiers in Neurorobotics, vol. 7, p. 21, 2013.
[11] A. McDermott, N. Deighton, and J. L. Walsh, "Machine learning for mid-infrared spectroscopy-based protein prediction in milk," Journal of Dairy Science, vol. 104, no. 2, pp. 1848–1858, 2021.
[12] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, no. 1, pp. 281–305, 2012.
[13] X. S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," Neural Computing and Applications, vol. 24, no. 1, pp. 169–174, 2014.
[14] A. H. Gandomi, X. S. Yang, and A. H. Alavi, "Cuckoo search algorithm: A metaheuristic approach to solve structural optimization problems," Engineering with Computers, vol. 29, no. 1, pp. 17–35, 2013.
[15] Y. Zhang and X. S. Yang, "Cross-disciplinary applications of cuckoo search algorithm," Archives of Computational Methods in Engineering, vol. 25, no. 4, pp. 1055–1075, 2018.
[16] X. S. Yang and S. Deb, "Cuckoo search via Lévy flights," in Proc. World Congr. Nature Biol. Inspired Comput. (NaBIC), 2009, pp. 210–214.
[17] Y. Zhang and X. S. Yang, "Cuckoo search for hyperparameter optimization in deep learning," Applied Soft Computing, vol. 94, p. 106450, 2020.
[18] B. Wahyudi, et al., "Parameter optimization of Cuckoo Search Algorithm for machine learning applications," Journal of Computational Optimization, vol. 17, no. 4, pp. 355–368, 2023.
[19] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," Analytical Chemistry, vol. 36, no. 8, pp. 1627–1639, 1964.
[20] S. Lohumi, H. Lee, M. S. Kim, J. Qin, and B. K. Cho, "Raman spectroscopy coupled with chemometrics for food authentication: A review," TrAC Trends in Analytical Chemistry, vol. 107, pp. 196–208, 2018.
[21] D. Hanganu and N. Chira, "Reassessing NMR peaks for dairy authentication," Journal of Dairy Science, vol. 104, no. 3, pp. 3275–3284, 2021.
[22] L. Y. Foo, et al., "Traceability and transparency in food fraud prevention using blockchain," Food Chemistry, vol. 372, p. 130988, 2021.
[23] A. Junaidi, et al., "Spektroskopi dan machine learning untuk autentikasi produk pangan lokal," ICEEI, vol. 1, no. 1, pp. 13–26, 2024.
[24] A. Sitorus, et al., "Rapid detection of coconut milk adulteration using FTIR spectroscopy and chemometric analysis," Elsevier, 2021.
[25] N. L. Yusof, et al., "Multivariate analysis in authenticating coconut-based beverages," Foods, vol. 12, no. 1, p. 199, 2021.
[26] Y. Zhang, et al., "Use of FTIR and chemometrics in plant-based milk authentication," Sensors, vol. 24, no. 1, p. 2362, 2022.
[27] N. Lestari, et al., "Remote sensing-based evaluation for supply chain fraud prevention," Remote Sensing, vol. 15, no. 18, p. 3764, 2025.
[28] A. Budi, et al., "Deteksi pemalsuan santan instan dengan spektroskopi FTIR," Food Chemistry Advances, vol. 3, p. 100021, 2025.
[29] G. Airlangga, "Enhancing medical diagnostics with ensemble machine learning: A comparative study," Jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK), vol. 9, no. 2, pp. 1117–1132, 2024.
[30] Y. Zhang and X. S. Yang, "Cuckoo search for business optimization application," Applied Intelligence, vol. 52, pp. 1820–1837, 2020.
[31] H. Wijaya *et al.*, "Application of FTIR spectroscopy for rapid and non-destructive analysis of adulteration in coconut milk," *J. Food Compos. Anal.*, vol. 97, p. 103768, 2021.
[32] R. J. Barnes *et al.*, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Appl. Spectrosc.*, vol. 43, no. 5, pp. 772–777, 1989.
[33] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York: Springer, 2013.
[34] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2020.
[35] P. Probst *et al.*, "Hyperparameters and tuning strategies for random forest," *WIREs Data Min. Knowl. Discov.*, vol. 9, no. 3, p. e1301, 2019.
[36] X. S. Yang, *Nature-Inspired Metaheuristic Algorithms*, 2nd ed. Frome, UK: Luniver Press, 2010.
[37] A. H. Gandomi and X. S. Yang, "Chaotic bat algorithm," *J. Comput. Sci.*, vol. 5, no. 2, pp. 224–232, 2014.
[38] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
[39] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.
[40] P. Virtanen *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, pp. 261–272, 2020.
[41] A. Rohman and Y. B. Che Man, "FTIR spectroscopy combined with chemometrics for analysis of lard adulteration in virgin coconut oil," *Food Anal. Methods*, vol. 5, no. 4, pp. 828–834, 2012.
[42] A. M. Marina *et al.*, "Monitoring virgin coconut oil adulteration by FTIR spectroscopy and chemometrics," *J. Food Sci. Technol.*, vol. 56, no. 4, pp. 2183–2192, 2019.
[43] P. M. Santos *et al.*, "Chemometrics and FTIR spectroscopy: A team to fight food fraud," *TrAC Trends Anal. Chem.*, vol. 127, p. 115877, 2020.
[44] O. I. Mba *et al.*, "Challenges in infrared spectroscopy for food authentication," *Crit. Rev. Food Sci. Nutr.*, vol. 61, no. 5, pp. 826–854, 2021.
[45] X. Zhang *et al.*, "Intelligent FTIR spectroscopy for food authentication: A tutorial review," *Anal. Chim. Acta*, vol. 1193, p. 339405, 2022.
[46] E. C. Y. Li-Chan, "The basics of infrared spectroscopy for food quality analysis," *Food Chem.*, vol. 217, pp. 342–353, 2017.
[47] A. Rohman *et al.*, "FTIR-based rapid detection of coconut milk adulteration: Recent advances," *J. Food Compos. Anal.*, vol. 108, p. 104411, 2022.
[48] R. Ríos-Reina *et al.*, "Preprocessing of FTIR spectra for food authentication: A comparative study," *Chemometr. Intell. Lab. Syst.*, vol. 212, p. 104287, 2021.
[49] D. Suhandy *et al.*, "Data fusion of FTIR and NIR spectroscopy for coconut products authentication," *Food Control*, vol. 145, p. 109432, 2023.
[50] J. Wang *et al.*, "1D-CNN for FTIR spectral classification of edible oils," *Spectrochim. Acta A*, vol. 285, p. 121891, 2023.
[51] A. M. Marina *et al.*, "FTIR spectral markers for traditional vs industrial coconut milk," *LWT-Food Sci. Technol.*, vol. 173, p. 114362, 2023.
[52] Y. B. Che Man *et al.*, "Chemometric analysis of coconut milk stabilizers using FTIR," *J. Food Sci.*, vol. 87, no. 5, pp. 1988–1999, 2022.
[53] T. Pirak *et al.*, "Geographical discrimination of Southeast Asian coconut milk by FTIR-chemometrics," *Food Chem.*, vol. 404, p. 134612, 2023.
[54] E. Hatzakis *et al.*, "NMR-FTIR combined approach for coconut lipid analysis," *Talanta*, vol. 253, p. 123943, 2023.
[55] X. Zhang *et al.*, "Real-time quality control of coconut milk production using portable FTIR," *J. Food Eng.*, vol. 337, p. 111234, 2023.