p-ISSN 2615-787X e-ISSN 2615-790X
Accredited by Directorate General of Higher Education, Research, and Technology, Republic of Indonesia, No. 225/E/KPT/2022

Tropical Animal Science Journal, September 2026, 49(5):406-417
DOI: https://doi.org/10.5398/tasj.2026.49.5.406
Available online at https://journal.ipb.ac.id/index.php/tasj

NIRS and Machine Learning Integration for Dynamic Monitoring of Nutritional Properties in Citronella Residues during Solid-State Fermentation

Samadi[1,2,*], E. Kaloudis[3], I. Wahyudi[1,2], S. Wajizah[1,2], & A. A. Munawar[4]
[1] Department of Animal Science, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[2] Research Center for Innovation and Feed Technology, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[3] Computer Simulation, Genomics and Data Analysis Laboratory, Department of Food Science and Nutrition, School of the Environment, University of the Aegean, Myrina 81400, Greece
[4] Department of Agricultural Engineering, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
*Corresponding author: samadi177@usk.ac.id
(Received 10-03-2026; Revised 20-05-2026; Accepted 22-05-2026)

ABSTRACT

Solid-state fermentation (SSF) can improve the nutritional value of agro-industrial residues by enriching microbial protein and modifying fiber structure. However, routine monitoring of these dynamic changes by conventional chemical methods is time-consuming. This study evaluated the feasibility of combining near-infrared spectroscopy (NIRS) with chemometrics and machine learning for rapid estimation of gross energy (GE), crude protein (CP), and crude fat (ether extract [EE]) in citronella residues during SSF. Citronella residue was fermented for 28 days under four white-rot fungal treatments and an uninoculated control (n = 60). NIR spectra (1000–2500 nm) were preprocessed using multiplicative scatter correction (MSC), and informative wavelengths were selected using the successive projections algorithm (SPA). Predictive models were developed using partial least squares regression (PLSR), support vector regression (SVR), and light gradient boosting machine (LightGBM), with a 3:1 split for calibration and prediction samples (45/15 samples). CP, the principal indicator of protein enrichment during SSF, can be effectively monitored using NIRS, with PLSR yielding the highest predictive accuracy among all models (R^2 = 0.72, RMSE = 0.29%, and MAE = 0.23%). The ability to monitor CP through NIR enabled the rapid evaluation of nutritional upgrading. GE showed moderate predictive performance, with the best results from LightGBM (R^2 = 0.55, RMSE = 192.34 J/g, and MAE = 146.02 J/g). The poor predictive capability of EE (best R^2 = 0.15) reflects the variability and low concentrations of EE, the weak and non-unique spectral signatures of EE, and strong spectral absorption in the sample matrix. Therefore, these data support the use of NIRS to rapidly evaluate CP yields and changes in the dynamic range of CP during SSF, but indicate that EE will continue to have limited value as a NIRS-based predictor in this highly complex fermented matrix.

Keywords: citronella residues; monitoring; NIRS; nutritional value; solid-state fermentation

INTRODUCTION

The conversion of biomass from agro-wastes into animal feed through bioconversion is a potential method for generating sustainable ruminant feed. Additionally, to improve sustainability in livestock production, this method also ensures long-term feed security while creating an environmentally friendly method of feeding animals (Ponnampalam *et al*., 2025). Citronella residue (CR) represents a significant portion of agro-industrial wastes that are readily available within many tropical production systems. It is estimated that over 99.2% of all waste produced in the citronella essential oil extraction industry is CR (Manurung *et al*., 2015). CR is primarily wasted due to its high level of lignin, which limits the nutritional value of the material as ruminant feed. Although it was found that ruminants may have limited intake when fed CR alone, there are still opportunities to use this material as a source of energy for ruminants. Solid-state fermentation (SSF) using filamentous fungi has recently become increasingly popular as a means to biologically modify the chemical composition of these materials. As a result of biological modification, SSF can reduce lignin content, increase microbial protein concentration, and decrease the presence of anti-nutritional compounds present in the substrate (Wang *et al*., 2023; Fan *et al*., 2024). As Heidari *et al*. (2022) stated, SSF of canola meal using *Pleurotus ostreatus* resulted in significantly decreased phytate concentrations (up to 98.8%), as well as an increase in crude protein concentration of 11%-18% compared to unfermented canola meal.

In ideal circumstances, SSF will also increase the nutritional quality of the feed materials. In an

SSF environment, organisms capable of producing exogenous enzyme production, i.e., cellulase and xylanase, may convert cellulose and hemicellulose-based oligo-saccharide polysaccharides into simpler oligo-saccharide chains for better accessibility by fiber-associated proteins (Chukwuma *et al*., 2020). The enzymatic activity in SSF processes occurs in a dynamic manner and is extremely sensitive to environmental variables, e.g., substrate variability, moisture distribution, temperature gradient differences, O_2 diffusion limitations, etc. (Bamidele *et al*., 2025). Therefore, Jiang *et al*. (2015a) asserted that to improve fermentation rate and prevent slow and/or poor fermentation rates, it is necessary to utilize effective online monitoring techniques, which will provide timely/real-time data for improving the consistency and quality of SSF processed products as value-added feedstuffs.

Near-infrared spectroscopy (NIRS) is a non-destructive analysis tool that has become increasingly popular across many different types of industries. Some examples include food (Jiang *et al*., 2023), agriculture (Munawar *et al*., 2024), pharmaceutical (Dégardin *et al*., 2016; Massei *et al*., 2025), and animal feed (Samadi *et al*., 2023, 2024, 2025). One reason why NIRS is so popular is that it offers a relatively fast and accurate way to get chemically relevant information when compared to other methods of collecting this type of data for evaluating biological processes. Using NIRS theoretically enables one to identify changes in the overtone and combination bands related to vibrations caused by stretching of the covalent bonds found within molecules, i.e., O-H, C-H, N-H, and S-H. These molecular vibrations are occurring in the near-infrared (NIR) range (Cen & He, 2007). Therefore, chemical information about biological systems can be obtained in the NIR region. During SSF, the substrate undergoes a very rapid biochemical transformation as the microorganisms break down the cellulose and hemicellulose present on the substrate. This leads to a significant increase in the moisture content, as well as the amount of microbial biomass produced and the level of protein and structural carbohydrates in the substrate. With each of these biochemical transformations, new functional groups are formed while others may be destroyed. Each of these biochemical transformations will alter the absorbance spectrum of the substrate in the NIR wavelength (Jiang *et al*., 2015b). Since there are differences in spectral characteristics for each of these biochemical transformations, spectral fingerprinting can be utilized to differentiate among the several levels of fermentation. Previous research has demonstrated that NIRS can be applied to measure moisture content and pH (Wahyudi *et al*., 2025), soluble protein and trypsin inhibitor content (Dai *et al*., 2023), and extent of fermentation (Jiang *et al*., 2015b) during SSF. In summary, these studies demonstrate that NIRS is a rapidly applicable, reliable, and versatile analytical tool that can provide data on both physical/chemical attributes as well as changes in anti-nutritional compounds throughout the SSF process.

Gross energy (GE) and crude protein (CP) and crude fat (ether extract or EE) are generally considered among the three most important of all of the nutrient parameters in evaluating ingredients of animal feeds; they provide the best general indication of the potential energy content of an ingredient to animals (Kil *et al*., 2013); indicate the bioavailability of amino acids and microbial nitrogen availability (Dias *et al*., 2025); and serve as the principal measure of the biochemical stability of fats in animal diets (Rolls, 2009). In recent years, there have been significant advancements in the use of NIRS technology, along with multivariate statistical techniques such as chemometrics and machine learning (ML), primarily in the field of animal feed and food science. For instance, Cavallini *et al*. (2025) indicated that a number of researchers have employed technologies such as those mentioned previously to develop "precision feeding" programs using a variety of fibrous agro-industrial by-product substrates, and consequently reducing the environmental impact associated with livestock production systems. Gasparini *et al*. (2026) argue that, while NIRS-based methodologies have been increasingly accepted due to their rapid determination of the chemical composition of food and feed ingredients at an industrial level, there is increasing acknowledgment of the relevance of NIRS-based methodologies for rapid assessment of the existence of contaminants and for identification of the origin and/or identity of individual feed ingredients within global supply chains. Therefore, although it is well established that NIRS-machine learning (NIRS-ML) can be successfully applied in a wide variety of applications beyond just predicting nutrients, this study aims to explore, for the first time, to the authors' knowledge, the capability of this methodology for the dynamic assessment of proximate nutritional component characteristics (i.e., GE, CP, and EE) during SSF of CR. Therefore, while the present study does not propose the introduction of NIRS-ML as a new paradigm for predicting proximate composition, we do intend to provide evidence (to our knowledge, for the first time) of its utility for dynamically measuring key proximate nutrient components (GE, CP, and EE) specifically in CR undergoing SSF and in a substrate/application context that has never previously been examined.

Although there is a potential for high levels of variability in the proximate nutritional values produced from SSF, this study aimed at developing and testing an NIRS-based predictive model that could predict changes in three key proximate nutritional components (gross energy [GE], crude protein [CP], and ether extract [EE]) as they occur during the SSF process. As the SSF process involves dynamic interactions among microbial populations, substrate materials, temperature, moisture, etc., it also introduces significant complexity into the data, potentially leading to redundant information in multidimensional NIR spectral datasets. Due to this level of complexity, variable selection methods are required to isolate wavelengths that contain the most useful information for predicting proximate nutrient content while minimizing interference with prediction performance.

To address these challenges, a successive projection algorithm (SPA) was applied to identify the appropriate wavelengths with the least multicollinearity prior to modeling. The three machine learning algorithms used to build the models are: light gradient boosting machine (LightGBM), partial least squares regression (PLSR), and support vector regression (SVR). These algorithms are used together because they have different advantages and disadvantages and have been demonstrated to be successful in chemometrics with small to medium sample sizes (n = 45 for calibration). PLSR is considered to be the traditional standard in NIRS analysis, as it addresses the problem of collinearity in the spectral data by reducing the dimensionality of the data into fewer principal components, which can perform adequately if there is a relatively simple linear relationship between the spectra and the measured property. SVR was used to address potential nonlinear relationships between spectra and the analyte, given the complexities of biochemical reactions during SSF. SVR has particular appeal for the large number of variables with few samples, since it uses both regularization and the "kernel trick" to reduce the risk of overfitting (Wahyudi *et al*., 2025). A further advantage of LightGBM compared to many other tree-based methods is both faster training times and lower memory usage. These advantages should be especially beneficial when you have multiple iterations of cross-validation and/or hyperparameter tuning. Due to the small sample sizes, alternatives such as deep neural networks were deliberately excluded from this comparison, as they would likely exhibit severe overfitting and poor generalization. This research is designed to test the effectiveness of three techniques (a linear latent variables technique (PLSR), a nonlinear regression technique (SVR), and a decision tree ensemble method (LightGBM) to identify which one can be most effective to follow the dynamic changes of substrate concentration during SSF fermentation by analyzing nutritional changes.

MATERIALS AND METHODS

Substrate and Spawns Preparation

Citronella residues (CR) from the citronella oil processing industry in Gayo Lues District of Aceh Province, Indonesia, were provided by local farmers. The CR was then cut into pieces that averaged about 3 cm in length. The CR was air-dried using a forced air drying oven at 60 °C for approximately 2 days. By doing so, the structure of the starches was preserved while reducing the moisture level to 10%-12% (Gao *et al*., 2024). The four white-rot fungi species tested in this project are: *Phanerochaete chrysosporium* (PCH), *Pleurotus ostreatus* (POS), *Trichoderma viride* (TV), and *Lentinula edodes* (LED). All the fungal strains studied here were acquired from the Indonesian Culture Collection (InaCC) Laboratory located within BRIN in Indonesia. Prior to fermentation, all fungi were activated and cultured on potato dextrose agar (PDA) plates, as described by Tuyen *et al*. (2013) at 24 °C until mycelia colonized the medium. Then, spawns were prepared by placing cultured agar (1 × 1 cm) into sterilized corn grain and then incubating at 24 °C until all grains were colonized by mycelia. Spawns were stored in a refrigerator (6 °C) for further cultivation.

Experimental Design and Solid-State Fermentation Procedure

The experiment employed a completely randomized design with 4 fungal treatments and one uninoculated control (WOI), with 3 replications. SSF was conducted on a pilot basis using Polyethylene bags. The Polyethylene bag for each treatment consisted of 447 g of CR, 30 g of molasses, and 100 g of corn bran. To achieve homogeneity, all ingredients were mixed together. The substrate was then inoculated with 50 g of spawn of each fungus (PCH, POS, TRV, and LED). While mixing, sterile water was added to the mixture in order to keep the total moisture content constant at 60%. SSF and WOI were both aerobically incubated at ambient temperature conditions for 28 days under natural fluctuation of temperature (±3 °C), approximately 37 °C.

Fermentation Periods and Sampling Strategy

The 28-day SSF was divided into 4 sequential 7-day phases that could be used to monitor the progression of nutrient availability throughout the SSF process. Each phase contained three sets of experiments: fungal treatment set-ups and control set-ups without inoculum; therefore, a total of 15 experimental units per time point. At each time point, an aseptic representative sample from each experimental unit was collected and subsequently evaluated using both NIR and chemical reference data. In total, 60 samples were collected at the four time points, providing a clear and complete picture of how nutrient content varied over the SSF process. The total number of samples (n=60) was determined by the SSF experiment design, which included 4 fermentation time points, 5 treatment conditions, and 3 biological replicates. As this number of samples is sufficient for preliminary proof-of-concept evaluations, however, due to having n=45 calibration samples, it can be stated that this dataset would be too limited to train multiple parameter models, particularly those classified as ensemble methods (i.e., LightGBM), as they will tend to overfit when trained on very small datasets. Therefore, results based upon model predictions should be viewed with extreme caution. Future studies will need to increase the number of calibration samples by increasing the number of SSF fermentation batches, adding treatments with fungi, or adding other potential substrate sources to obtain a larger, more robust calibration sample set.

NIR Spectra Acquisition

The NIR spectrum of each of the SSF samples was collected immediately upon collection of each SSF sample at every sampling time point. A sample holder containing approximately 5 g of each SSF sample was then gently pressed down onto a flat area inside the holder to create a level plane for spectroscopy. NIR spectra were measured using a NIRFlex N-500 spectrometer (Büchi, Flawil, Switzerland), operated in reflection mode, collecting

absorbance (log 1/R) across wavelengths from 1000–2500 nm. For each sample spectrum, three measurements were made by averaging thirty-two sequential measurements.

Validation of the spectrometer's performance and calibration transfer: the quality of the data produced by the spectrometer was validated prior to each sampling event using an internal polystyrene reference standard provided by the manufacturer, which verified that the wavelengths used during measurement were accurate and that the signal-to-noise ratio did not vary significantly. Since no calibration transfer process was conducted for comparison between the multiple sample events (all samples were measured on one spectrometer) or between the different instruments, since all measurements occurred during a single experiment, there is no indication that the models generated in this project are generalizable to any other type of spectrometer, therefore further studies may be required to validate these models on a wider range of instrumentation.

Chemical Analyses

The samples were analyzed using reference chemicals to determine their gross energy (GE), crude protein (CP), and ether extract (EE) content. The GE values were obtained through Bomb calorimetry. The CP (N x 6.25) and EE values were measured in accordance with AOAC standardized methodologies for Kjeldahl and Soxhlet extraction, respectively (AOAC, 2016).

Spectra Preprocessing and Wavelengths Selection

Using the successive projections algorithm (SPA), the most important wavelength(s) were identified using the NIR spectra data. The success of SPA in identifying the most important wavelengths is especially significant when dealing with high-dimensional spectral data (like NIR spectra), which generally have far more wavelengths than sample points (He *et al*., 2025a). SPA utilizes an iterative forward variable selection technique, which minimizes the redundancy of the selected variables while maximizing their ability to contain as much predictive model information as possible (He *et al*., 2025a). By eliminating multicollinearity among all selected variables and choosing only those most effective at creating a reliable predictive model, calibration models can be made as effective as possible.

Before selecting the best SPA for our spectral data, the authors first applied multiplicative scatter correction (MSC) to all of it. This was done because MSC is a widely used preprocessing method for NIR data that removes the effects of sample physical properties, thereby reducing differences in the degree of light scattered from different areas of the same sample (Li *et al*., 2019); thereby improving the relationship between spectra and reference values. The authors did not limit the number of wavelengths selected by the SPA. The authors allowed SPA to internally select its own wavelength numbers. Therefore, for every possible number of selected variables (1-50; there are about 1500 wavelengths in our full spectrum), SPA will perform a forward selection operation and evaluate each of these sets using the root mean square error of cross-validation (RMSECV) on our calibration set (n=45). Our final decision on the number of wavelengths to use was based on the lowest RMSECV. Additionally, a parsimony constraint was also used to help prevent overfitting. After optimizing the number of wavelengths, the SPA selected 35 wavelengths for gross energy (GE), 14 wavelengths for crude protein (CP), and 14 wavelengths for crude fat (EE) for model building/prediction.

Machine Learning Modeling and Evaluation

The development process of the NIRS-ML model in this paper is illustrated with a flow diagram shown in Figure 1. Before applying ML models on the spectra selected by SPA, these wavelengths were first divided into two groups, namely calibration group (70% of samples, n=45) and prediction group (30% of samples, n=15) using a random sampling technique called "stratified sampling" to keep reference value distributions intact. The number of samples used for calibration versus prediction groups was selected with consideration of prior research. That research indicated that a large calibration group was needed to account for spectral differences while preserving a valid set of samples for validation (Samadi *et al*., 2023; Fodor *et al*., 2024). Since Table 1 shows similarities in mean and standard deviation values between the calibration and prediction groups for all analytes, it can be inferred that the random stratification resulted in a reasonable sample of the total population (n = 60), given the small overall sample size. A potential limitation of the current work is that a completely separate, independent prediction test group was established, yet with so few samples (n = 15), the predictive performance estimates may not accurately reflect generalization to other production runs or operating conditions. Spectral data were considered predictor variables while GE, CP, and EE were considered response variables. Two additional machine learning algorithms were tested and compared to partial least squares regression (PLSR) as follows. In addition to developing PLSR through the process of maximizing the covariance between the spectral data and target parameters (i.e., GE, CP, and EE) using latent variables (LVs), the authors also developed two other machine learning models (support vector regression [SVR], and light gradient boosting machine [LightGBM]). The SVR model was developed using a radial basis function (RBF) kernel to potentially identify nonlinear relationships between individual features within the higher dimensional spectral space (He *et al*., 2025b; Wahyudi *et al*., 2025). Because of the small training set size (n = 45), there is also significant concern about overfitting with LightGBM, as tree-based ensemble models tend to fit noise in smaller sample sizes.

The parameters for both SVR and LightGBM were determined through a ten-fold cross-validation method using a grid search in order to find the optimal hyperparameters. In this regard, there were three variables that were used to create the grid for the SVR. These variables were *C* (0.1, 1, 10, 100), *gamma* (0.001, 0.01, 0.1), and *epsilon* (0.01, 0.1, 0.2), while the LightGBM grid included *num_leaves* (15–63), *learning_rate* (0.01–0.1), *n_estimators* (100–500), *max_depth* (−1 to 10), and *min_*

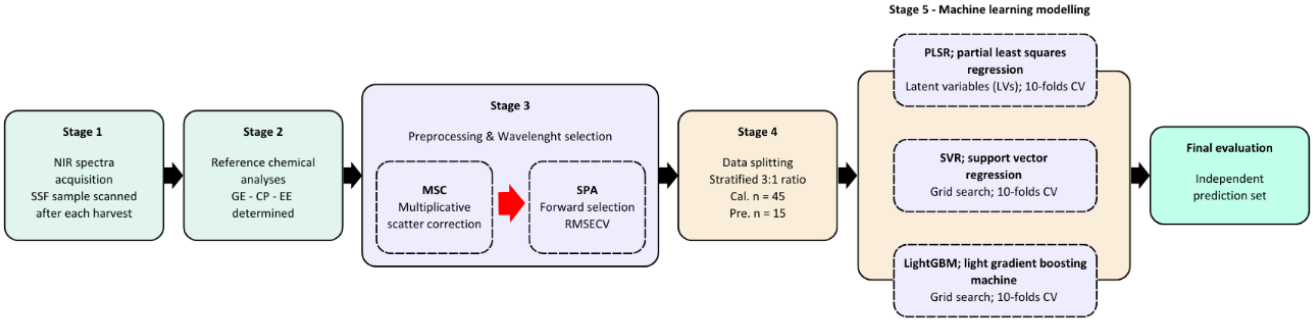

Figure 1. A flow chart of the NIRS-ML model development process

Table 1. Descriptive statistics of actual measurements of nutritional content from whole fermented citronella residues in the dataset

Statistics	Gross energy (J/g)		Crude protein (%)		Crude fat (%)	
	Cal	Pred	Cal	Pred	Cal	Pred
n	45	15	45	15	45	15
Min	16221	16328	6.95	7.15	0.2	0.88
Max	17337	17252	9.01	8.99	4.39	4.44
Mean	16785.22	16797.40	8.00	8.00	2.38	2.50
SD	281.76	295.66	0.57	0.56	0.96	0.99
Var	79387.27	87417.69	0.32	0.32	0.92	0.98

Note: Cal, calibration; Pred, prediction; n, number of sample datasets; Min, minimum; Max, maximum; SD, standard deviation; Var, variance.

child_samples (10–30). There were five variables that were used to create the grid for LightGBM. They were the *num_leaves* (15–63), *learning_rate* (0.01–0.1), *n_estimators* (100–500), *max_depth* (−1 to 10), and *min_child_samples* (10–30). Ultimately, the best combination of the hyperparameters for the best fit model was determined based on the model with the smallest root mean squared error (RMSE) as measured against all combinations tested. All models were then run against an independent test set to evaluate their predictive ability.

The model's performance was assessed with the use of calibration and prediction coefficient of determination (R^2_c and R^2_p) along with associated root mean square error ($RMSE_c$ and $RMSE_p$) in addition to R^2 and RMSE values above 0.80 for the former, or less than the reference standard deviation (SD) for the latter as an indicator of acceptable calibration and predictability respectively (Samadi *et al*., 2023; Munawar *et al*., 2024). Additionally, the model's performance was also assessed using two other metrics. These were the mean absolute error (MAE), which calculates the average of the absolute differences between observed and forecasted outcomes. MAE provides a very easy-to-interpret measure of how well a regression model predicts the outcome variable. Overall, the smaller the MAE, the better the predictive capability of the model (Khadem *et al*., 2024).

There are no universally accepted thresholds for "good" or acceptable levels of accuracy in predicting performance with NIRS, since the criteria for evaluating prediction performance vary depending on the analyte being measured, the complexity of the sample matrix, and the purpose for which predictions will be used. The chemometric and food/feed NIRS literature provides some general guidelines that have been adopted here as references for evaluating the performance of models developed in this study: An R^2 value of at least 0.9 is typically considered to represent an excellent level of predictive ability and to be suitable for use as part of a routine quality control program (García-García *et al*., 2022). Useable predictive ability exists when $0.7 \leq R^2 < 0.9$ and if the root mean square error (RMSE) is also at a level that makes it useful for screening/monitoring processes (Nicolai *et al*., 2007; Pasquini, 2018). Predictive ability is considered poor when $R^2 < 0.5$, and the model should not be relied upon to identify the relative concentrations of samples to within an order of magnitude (Saeys *et al*., 2005; Fodor *et al*., 2024).

All data analysis regarding spectroscopy and machine learning models, as well as graphically displaying results, was completed in a Google Colab (12 GB RAM) environment running Python 3.1.3. PLSR was implemented using the *PLSRegression* class from the *cross_decomposition* module of the scikit-learn library, while SVR was constructed using the *SVR* class from the *sklearn.svm* module. The LightGBM model was built using the *LGBMRegressor* class from the *lightgbm* library. For the hyperparameters in SVR and LightGBM, an exhaustive grid search was performed by utilizing *GridSearchCV* from *sklearn.model_selection* module. Performance of each model was measured by calculating the RMSE, R^2 values, and MAE, which were all calculated using the *sklearn.metrics* module. Tabular results were compiled, and data manipulation was accomplished through the utilization of the *pandas* library. Graphical visualizations were also generated using the *matplotlib* library.

RESULTS

The results section provides information on gross energy (GE), crude protein (CP), and crude fat (EE) using a proposed workflow based on NIRS. Initially,

the subsets used to calibrate the model and predict the data are described (Table 1). Following this description of subsets, the wavelength selected from MSC pre-processed data after selecting principal variables by SPA (Tables 2-3; Figures 2-4) and subsequently the predictive capabilities of models derived with PLSR, SVR and LightGBM using R^2, root mean square error (RMSE), and mean absolute error (MAE) on an independent test dataset (Tables 3; Figures 5-8).

Dataset Characteristics and Implications for Calibration

The descriptive statistics (Table 1) suggest that the calibration (n=45) and prediction (n=15) subsets exhibit the same central tendency and dispersion for each of the three response variables, thus supporting the use of the current stratification ratio of 3:1 as an initial feasibility evaluation. GE has a narrower range and similar means between subsets (16,785.22 vs. 16,797.40 J/g) than CP, and also exhibits identical mean values (8.00%), as well as very similar dispersion (SD = 0.57 vs. 0.56). EE has the greatest relative spread (0.20%-4.39% in calibration; 0.88%-4.44% in prediction; Table 1). The fact that one sample holdout set allows for an easy distinction to be made regarding training (calibration) vs. test (independent prediction), however, validation stability with respect to NIRS chemometrics will depend very heavily upon whether the data used to predict represents the same type of usage as originally planned.

Variable Selection and Spectral Interpretation

After the MSC process was completed on all the samples, the SPA was used to reduce the number of wavelength dimensions from 1557 to 35 (GE) and 14 (CP and EE), as shown in Table 2. The spatial distribution of selected wavelengths for each analyte is shown in Figures 2–4. The selected wavelengths were not uniformly distributed across the spectral range. For all

Table 2. NIR wavelengths selected by the successive projections algorithm after multiplicative scatter correction preprocessing

Parameters	n	NIR Selected Wavelength (nm)
Gross energy (J/g)	35	1192.1, 1231.7, 1186.1, 1242.9, 1185, 1189.9, 1185.5, 1188.8, 1229.4, 1277.2, 1190.4, 1191.5, 1193.7, 1191, 1228.2, 1187.7, 1184.4, 1187.1, 2051.2, 2046.4, 1225.3, 1238.2, 1024.8, 1188.2, 1237, 1183.9, 1020, 1022, 1444.4, 1233.5, 1186.6, 1182.8, 1193.2, 1235.8, 1189.3
Crude protein (%)	14	1191, 1186.6, 2178.8, 1568.5, 1187.7, 1185.5, 1995.9, 1361.7, 1183.9, 2242.8, 2044.7, 1228.2, 1231.7, 1237.6,
Crude fat (%)	14	1726.2, 1186.6, 1187.1, 1225.3, 1491.8, 1189.9, 1995.9, 1228.2, 1004.9, 2162.4, 1237.6, 2240.9, 2044.7, 1358.2

Note: n, number of selected wavelengths.

Table 3. Comparison of machine learning models based on SPA-Selected NIR wavelengths after multiplicative scatter correction

Parameters	Model	Best parameters	Calibration (n=45)			Validation (n=15)		
			R^2	RMSE	MAE	R^2	RMSE	MAE
Gross energy (J/g)	PLSR	LVs: 2	0.11	263.33	225.79	0.17	260.85	213.28
	SVR	C: 10, epsilon: 0.2, gamma: 0.1	0.71	150.90	102.83	0.51	200.26	155.07
	LightGBM	learning_rate: 0.1, max_depth: -1, min_child_ samples: 10, n_estimators: 500, num_leaves: 15	0.82	118.99	84.37	0.55	192.34	146.02
Crude protein (%)	PLSR	LVs: 10	0.58	0.36	0.29	0.72	0.29	0.23
	SVR	C: 100, epsilon: 0.2, gamma: 0.001	0.45	0.42	0.32	0.53	0.37	0.31
	LightGBM	learning_rate: 0.01, max_depth: -1, min_child_ samples: 10, n_estimators: 200, num_leaves: 15	0.54	0.38	0.30	0.57	0.36	0.32
Crude fat (%)	PLSR	LVs: 6	0.39	0.74	0.59	0.15	0.88	0.67
	SVR	C: 0.1, epsilon: 0.01, gamma: scale	0.14	0.88	0.72	-0.05	0.98	0.80
	LightGBM	learning_rate: 0.05, max_depth: -1, min_child_ samples: 20, n_estimators: 100, num_leaves: 15	0.20	0.85	0.70	-0.06	0.98	0.76

Note: PLSR, partial least squares regression; SVR, support vector regression; LightGBM, light gradient boosting machine; R^2, coefficient of determination, RMSE, root mean square error; and MAE, mean absolute error.

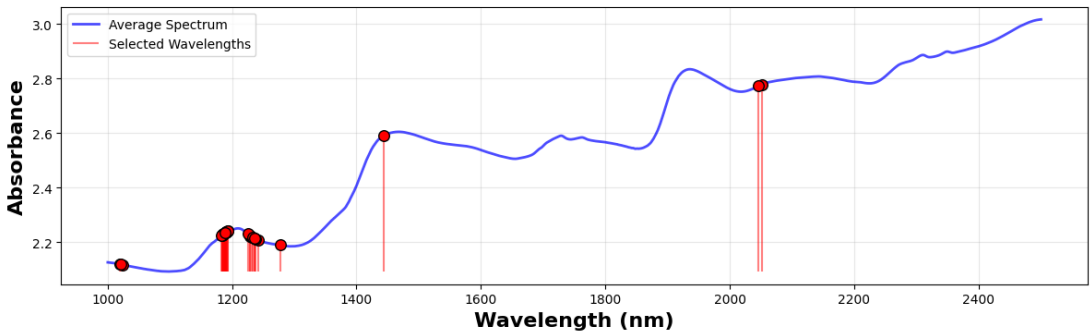

Figure 2. Distribution of the selected wavelengths (nm) used to predict gross energy. The blue line shows the average NIR absorbance spectrum of the samples. The red markers indicate the wavelengths that the SPA model identified as important spectral variables.

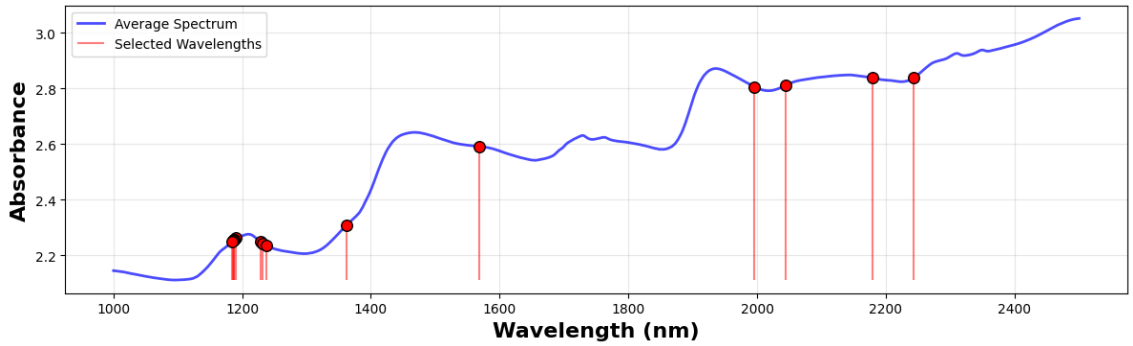

Figure 3. Distribution of the selected wavelengths (nm) used to predict crude protein. The blue line shows the average NIR absorbance spectrum of the samples. The red markers indicate the wavelengths that the SPA model identified as important spectral variables.

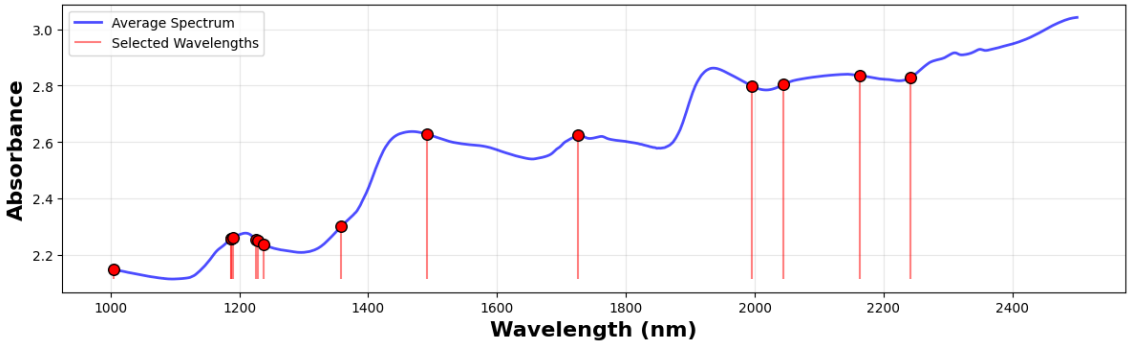

Figure 4. Distribution of the selected wavelengths (nm) used to predict crude fat. The blue line shows the average NIR absorbance spectrum of the samples. The red markers indicate the wavelengths that the SPA model identified as important spectral variables.

three analytes, a cluster of wavelengths was identified within the 1180–1245 nm region. Beyond this shared region, analyte-specific wavelengths were also selected: GE includes additional variables around 1020-1025 nm, 1444 nm, and 2046-2051 nm (Table 2, Figure 2), CP includes variables near 1361.7, 1568.5, 1995.9, 2044.7, 2178.8, and 2242.8 nm (Table 2, Figure 3), and EE includes variables such as 1491.8, 1726.2, 1995.9, 2044.7, 2162.4, and 2240.9 nm (Table 2, Figure 4). Although band assignment in complex SSF matrices should remain cautious, the observed differentiation is consistent with the concept that SSF drives biochemical transformations (e.g., lignocellulose degradation, microbial biomass synthesis, changes in protein/carbohydrate fractions).

Predictive Performance for Gross Energy

Model performance for GE is summarized in Table 3 and visualized in Figure 5. The nonlinear methods outperform PLSR on the independent prediction set: LightGBM achieves the highest validation performance (R^2 = 0.55; RMSE = 192.34 J/g; MAE = 146.02 J/g), followed closely by SVR (R^2 = 0.51; RMSE = 200.26 J/g; MAE = 155.07 J/g), whereas PLSR shows substantially weaker performance (R^2 = 0.17; RMSE = 260.85 J/g; MAE = 213.28 J/g). The scatterplots (Figure 5) reflect this ranking, with tighter alignment to the identity line for SVR and LightGBM than for PLSR, although dispersion remains evident. A divergence between calibration and validation performance was observed for LightGBM (R^2_c = 0.82 vs. R^2_p = 0.55; Table 3). Relative to the natural variability of GE in the prediction set (SD = 295.66 J/g), the lowest RMSE achieved (192.34 J/g for LightGBM) corresponds to approximately 65% of one standard deviation. The MAE summary in Figure 8 provides a consistent picture, with GE error levels intermediate between CP and EE across models.

Predictive Performance for Crude Protein

The CP model exhibited the best predictive performance among the models tested, based on R^2 and error metrics (RMSE, MAE), compared to SVR and LightGBM. In addition, a comparison of the R^2 values indicates that PLSR was superior (R^2 = 0.72; Table 3, Figure 6) when compared to both LightGBM (R^2 = 0.57) and SVR (R^2 = 0.53). Scatter plots using the CP model were generated (Figure 6) and demonstrate greater correlation between predicted and actual values for PLSR when compared to the two alternative models. As an additional measure, the standard deviation of predictions from CP (SD = 0.56) indicates that the $RMSE_p$ (0.29) represents less than one-half of the total data set variability.

Predictive Performance for Crude Fat

In contrast, EE prediction was unsatisfactory for all tested algorithms (Table 3, Figures 7-8). The best validation R^2 is low (PLSR, R^2 = 0.15), while SVR and LightGBM yield negative validation R^2 values (-0.05 and -0.06), implying performance worse than predicting

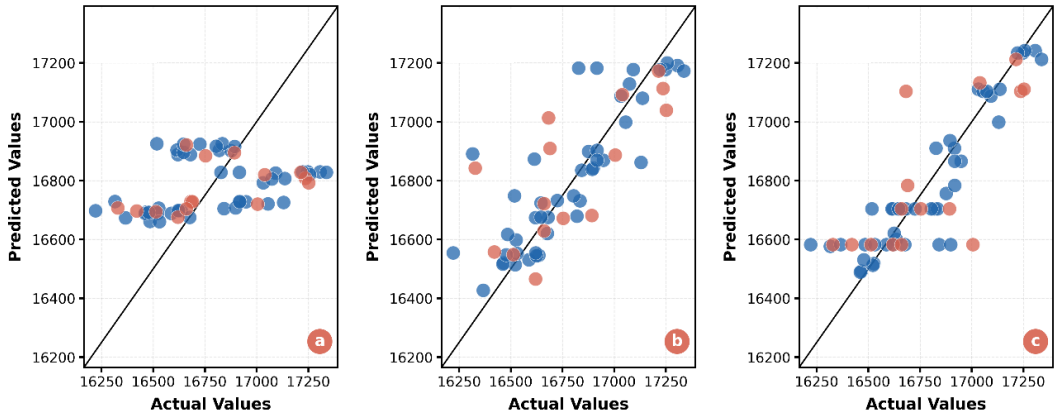

Figure 5. Scatter plots illustrating the relationship between measured and predicted gross energy (J/g) values for calibration (Cal) and validation (Val) datasets using (a) partial least squares regression (PLSR), (b) support vector regression (SVR), and (c) light gradient boosting machine (LightGBM) models. The diagonal line represents the ideal 1:1 relationship between measured and predicted values. ● Cal; ● Val.

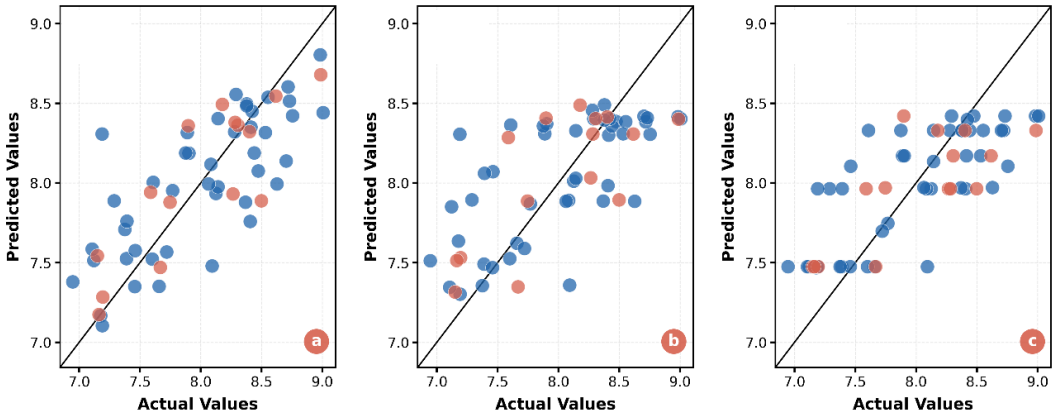

Figure 6. Scatter plots illustrating the relationship between measured and predicted crude protein (%) values for calibration (Cal) and validation (Val) datasets using (a) partial least squares regression (PLSR), (b) support vector regression (SVR), and (c) light gradient boosting machine (LightGBM) models. The diagonal line represents the ideal 1:1 relationship between measured and predicted values. ● Cal; ● Val.

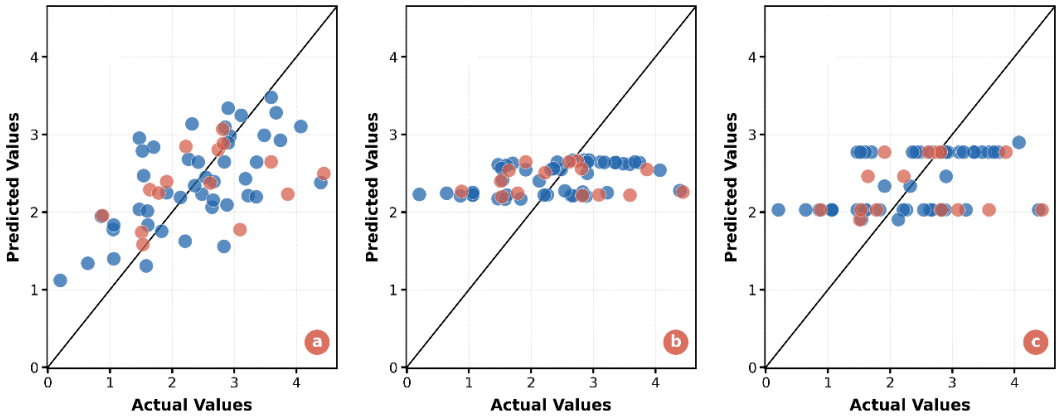

Figure 7. Scatter plots illustrating the relationship between measured and predicted crude fat (%) values for calibration (Cal) and validation (Val) datasets using (a) partial least squares regression (PLSR), (b) support vector regression (SVR), and (c) light gradient boosting machine (LightGBM) models. The diagonal line represents the ideal 1:1 relationship between measured and predicted values. ● Cal; ● Val.

the mean. This weak relationship is evident in Figure 7, where measured versus predicted values show wide scatter for all models, and in Figure 8, where EE MAE values remain the highest across targets. The error-to-dispersion comparison further clarifies the limitation: EE SD in the prediction set is 0.99 (Table 1), while $RMSE_p$ values range from 0.88 to 0.98 (Table 3), indicating that the prediction error approaches the full scale of one SD.

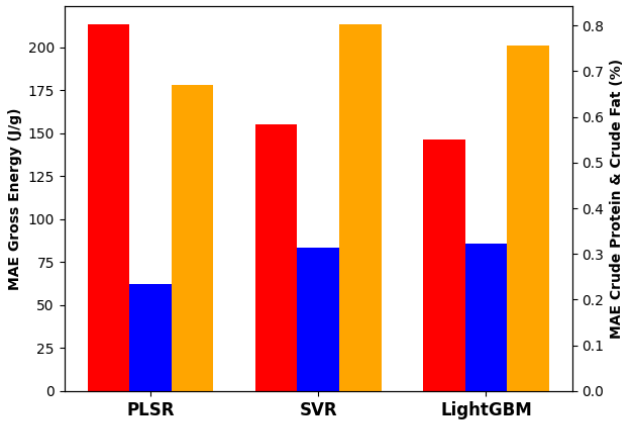

Figure 8. Comparison of prediction performance among partial least squares regression (PLSR), support vector regression (SVR), and light gradient boosting machine (LightGBM) models based on mean absolute error (MAE). The left y-axis corresponds to MAE values for gross energy (J/g), while the right y-axis represents MAE values for crude protein (%) and crude fat (%). Gross Energy; Crude Protein; Crude Fat.

DISCUSSION

The higher-order value of creating fast, reliable monitoring systems for SSF-produced feed ingredients is illustrated by the increased reliance on agro-industrial residues in animal feed production. Such materials exhibit natural variability in composition and, as seen in the recent cases cited in Gasparini *et al*. (2024), this variability may pose feed safety concerns due to the diversity of supply chains they originate from. As such, NIRS-based screening methods that track critical nutritional components during fermentation timelines at reasonable speeds (as shown to be possible with moderate reliability for CP and GE in this study) would provide an important first level of monitoring capability. The collected data collectively support a feasible analytical determination; however, this depends on the parameter analyzed. CP provides the highest degree of reliable analytical prediction (PLSR analysis, Tables 1-3, Figures 5-8); GE demonstrates some ability to predict moderately well using nonlinear models (LightGBM/ SVR, Table 3, Figure 5); and EE does not demonstrate sufficient ability to predict accurately (Table 3, Figure 7). The grouped MAE comparison in Figure 8 succinctly summarizes this hierarchy performance.

The use of four 7-day cycles (over 28 days) using SSF conditions in addition to the presence of replicates suggests that NIRS, coupled with chemometric/machine learning analysis, has potential as a fast-screening method to help monitor production processes specifically when nutritional enhancements are related to proteins (where predictions have been shown to be most reliable) (Jiang *et al*., 2015a; Dai *et al*., 2023). As the study did not provide empirical data to show generalization of the results to previously unseen periods of fermentation or novel uses of fungal treatments (i.e., time-hold-out validation or treatment holdout validation); the term "dynamic monitoring" should be defined as predictive models based on sampling at various points over time and/or treatment conditions as opposed to validated deployment under completely new periods/treatment (Ezenarro & Schorn-García, 2025).

Results from this study need to be understood within the context of several methodological limitations that reduce both the generalizability and reliability of the findings. As a limitation, the predictive accuracy of the models developed was based on a single holdout methodology (the 3:1 split shown in Table 1). Although this provides a starting point for estimating the potential predictive value of these models with respect to an independent test data set, this approach also does not provide estimates of predictive accuracy variability associated with the use of different splits or predictability when using multiple independent fermentation batches, which has become recognized as best practice in NIRS chemometric validation (Ezenarro & Schorn-García, 2025). Second, samples were generated under a structured SSF design (fermentation period × fungal treatment). Consequently, the reported prediction performance should be viewed as feasible within the sampled domain rather than evidence of universal generalization to unseen operating conditions, time periods, or fungal treatments. Third, the risk of overfitting was a specific concern for LightGBM given the small calibration set (n = 45). Several measures were taken to mitigate this: (i) a conservative hyperparameter grid was used, favouring shallow trees (*num_leaves*: 15-63, *max_depth*: -1 to 10) and a minimum child sample constraint (*min_child_samples*: 10-30) to limit model complexity; (ii) hyperparameter selection was based on 10-fold cross-validation RMSE rather than training error; and (iii) the *num_leaves* and *min_child_samples* parameters were explicitly constrained to prevent leaf-level memorisation of noise. Despite these measures, the calibration-prediction gap for GE (R^2_c = 0.82 vs. R^2_p = 0.55) indicates that overfitting was not fully eliminated. Accordingly, GE prediction should be framed as moderately informative for screening/process support within the present dataset rather than high-precision routine quantification. Finally, crude fat (EE) could not be predicted reliably in the current configuration, as indicated by low or negative validation R^2 values and prediction errors that remain high relative to the natural dispersion of EE in the dataset (Tables 1 and 3, Figures 7-8).

The limitations should be identified and explicitly communicated for each specific analytical component, with the understanding that they represent analyte-specific constraints associated with the current dataset and modeling configuration. Moreover, the structured nature of the data collected (i.e., samples were non-independent; rather, samples were nested within a defined experimental configuration) will limit the ability to generalize findings based upon sample number; therefore, it is likely that the effective size of the calibration set will be less than the total number of samples analyzed. Furthermore, hyperparameter optimization was performed with a single 10-fold cross-validation run using the full calibration set. Optimization via repeated or nested cross-validation runs could yield

more stable estimates of model performance and reduced sensitivity to partitioned subsets of the calibration data (Jalal *et al*., 2025).

While there are certainly some constraints associated with the research that was conducted here, this work is consistent with other reports concerning the use of spectroscopy as a monitoring tool for SSF processes by applying chemometrics. More specifically, FT-NIR has been demonstrated as an effective method to track key SSF process parameters, including the alcohol content and residual glucose (Jiang *et al*., 2018), which provides considerable justification for the use of NIRS as a monitoring tool during the SSF process independent of differing targeted analytical parameters. Therefore, this project may be viewed as an expansion of the monitoring process variables to include proximate nutrients. There was strong evidence for predicting crude protein content and moderate evidence for predicting total energy content, given the experimental parameters. The superior performance of nonlinear statistical methods demonstrates that the relationship between NIR spectral information and GE content in the complex SSF environment studied herein does not follow a strictly linear model. This outcome is also consistent with biological expectations, since GE content reflects the cumulative contributions of multiple biochemically distinct compounds (e.g., protein, fat, dietary fiber, residual carbohydrate) that vary independently in their rates and patterns of degradation throughout fermentation. Moreover, it is likely that overlapping absorption bands and matrix effects lead to nonlinear relationships among NIR spectral features that cannot be captured by linear latent variable models such as PLSR. Nevertheless, the calibration-prediction gap observed for LightGBM (R^2_c = 0.82 vs. R^2_p = 0.55) indicates that the calibration set was overly fit to the calibration data; a problem that occurs when flexible tree-based models (e.g., LightGBM) are employed with limited sample sizes (i.e., n = 45). Hence, while GE content prediction could yield reasonable results for screening ($RMSE_p \approx 0.65 \times SD$), the developed models should not be used for quantitative accuracy assessment without additional independent validation.

The practical benefit of the successful prediction of CP (R^2_p = 0.72; RMSE = 0.29%) will be very useful in practice. CP is one of the primary specifications used by commercial feed manufacturers to determine how much to pay for their ingredients and to develop rations. Traditional Kjeldahl chemical methods require approximately 4-6 hours to obtain results, plus additional time for sample preparation. However, using NIRS to predict crude protein content, with the developed PLSR model, estimates can be obtained immediately after measurement. Therefore, this rapidity allows the operator(s) to make adjustments to fermentation time, aeration rate, or inoculum quantity during processing that would otherwise occur at the end of processing, prior to product completion, thereby reducing variability from batch to batch. Additionally, NIRS provides small-scale or resource-limited feed manufacturers (including many feed producers in tropical areas where large quantities of citronella residue are readily available) with a cost-effective alternative to expensive, laboratory-based wet-chemical analytical techniques. Although the prediction accuracy for GE (R^2_p = 0.55; RMSEP = 192.34 J/g) does not meet the quality control thresholds ($R^2_p \geq 0.90$), it may still serve as a screening tool. For instance, if there are large deviations from predicted GE values during fermentation monitoring, this would indicate abnormal substrate degradation and/or incomplete fermentation. As such, targeted chemical analyses could then be performed on these suspect lots. It appears that the better-performing nonlinear models (LightGBM, SVR) capture complex relationships among the components (carbohydrate, protein, and fat) of biomass. However, the calibration prediction gap observed with LightGBM (R^2_c = 0.82; R^2_p = 0.55) suggests that current model complexity has yet to achieve full generality; therefore, GE predictions should be viewed cautiously and considered only as trending indicators.

In addition to being theoretically important, the inability to predict EE (negative R^2_p for SVR and LightGBM) provides practical information about how far NIRS can go with this matrix and compound. The inability to predict EE can be explained by several interacting chemical and spectroscopic variables. First, in the present data set, EE concentrations were very low and highly variable (range = 0.20%-4.39% in calibration; see Table 1), reducing the SVR available for model development. Lipid bands associated with CH-stretching overtone transitions from 1704-1780 nm, and CH-combination band transitions from 2300-2370 nm (Liu *et al*., 2015) are responsible for most of the absorption due to lipids in NIR spectra. However, in the more complex matrices of fermented substrates like those used here, these lipid bands are much weaker than the strong absorptions of water (OH, ~1450 & ~1940 nm), protein (NH, ~2050 & ~2180 nm), and structural carbohydrate molecules. These relatively large absorptivities will have masked the fat-related spectral signals. Additionally, the biochemical heterogeneity resulting from SSF processes, including partial breakdown of cell walls and redistribution of lipid fractions, is expected to increase spectral complexity and reduce the consistency of fat-related absorption patterns in samples. Manufacturers interested in using NIRS to measure fat levels in fermented citronella residue would need to supplement the substrate with sufficient fat (>5% EE) to obtain a measurable response or use an alternate method to rapidly measure fat content in their products, such as mid-infrared (MIR) spectroscopy. MIR spectroscopy provides significantly stronger, more selective lipid absorption bands than NIRS and has enabled improved predictions of fat content in more complex biological matrices (Lozano *et al*., 2017). From a fermentation monitoring perspective, EE is not a reliable process indicator under current conditions; moisture, pH, or CP are more appropriate targets.

CONCLUSION

This study confirmed that NIRS coupled with both chemometrics and machine learning can predict certain nutritional characteristics in citronella residue undergoing solid-state fermentation. The most reliable results were obtained by estimating CP utilizing PLSR. Gross energy (GE) was estimated at low reliability but with good

enough values to provide only approximate trending estimates. LightGBM had the best predictive performance for all models tested on GE. Crude fat (EE), however, could not be accurately predicted using this technique under these data sets and model configurations. While there is obvious value in NIRS technology as a quick screening method for tracking changes in crude protein levels throughout solid-state fermentation, it is also important to understand that the utility of NIRS has limitations; i.e., predictions for GE are likely only useful for general trending purposes and should not be taken as exact quantitative estimates. Additionally, the lack of reliable EE estimation means that no recommendations for use can be made without either advancements in methodology or additional calibration samples.

CONFLICT OF INTEREST

The authors declare that there are no known financial or personal conflicts of interest that could have influenced the findings presented in this study.

ACKNOWLEDGEMENT

The authors express their gratitude to Universitas Syiah Kuala for supporting this research through the H-Index Scheme in 2025, under contract number 327/UN11.L1/PG.01.03/14641-PTNBH/2025.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

The authors state that no generative AI or AI-assisted technologies were used in the writing of this manuscript.

REFERENCES

AOAC. (2016). Official Methods of Analysis of the Association of Official Analytical Chemists (20th ed.). AOAC Inc.

Bamidele, M. O., Bamikale, M. B., Cárdenas-Hernández, E., Bamidele, M. A., Castillo-Olvera, G., Sandoval-Cortes, J., & Aguilar, C. N. (2025). Bioengineering in Solid-State Fermentation for next sustainable food bioprocessing. Next Sustainability, 6, 100105. https://doi.org/10.1016/j.nxsust.2025.100105

Cavallini, D., Lamanna, M., Colleluori, R., Silvestrelli, S., Ghiaccio, F., Buonaiuto, G., & Formigoni, A. (2025). The use of rumen-protected amino acids and fibrous by-products can increase the sustainability of milk production. Frontiers in Veterinary Science, 12, 1588425. https://doi.org/10.3389/fvets.2025.1588425

Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. Trends in Food Science & Technology, 18(2), 72–83. https://doi.org/10.1016/j.tifs.2006.09.003

Chukwuma, O. B., Rafatullah, M., Tajarudin, H. A., & Ismail, N. (2020). Lignocellulolytic enzymes in biotechnological and industrial processes: A review. Sustainability, 12(18), 1–31. https://doi.org/10.3390/su12187282

Dai, C., Xu, X., Huang, W., Yan, P., Hou, Y., He, R., & Ma, H. (2023). Monitoring of critical parameters in thermophilic solid-state fermentation process of soybean meal using NIR spectroscopy and chemometrics. Journal of Food Measurement and Characterization, 17(1), 576–585. https://doi.org/10.1007/s11694-022-01628-3

Dégardin, K., Guillemain, A., Guerreiro, N. V., & Roggo, Y. (2016). Near infrared spectroscopy for counterfeit detection using a large database of pharmaceutical tablets. Journal of Pharmaceutical and Biomedical Analysis, 128, 89–97. https://doi.org/10.1016/j.jpba.2016.05.004

Dias, A. V. L., Granja-Salcedo, Y. T., Messana, J. D., Camargo, K. D. V., Alves, K. L. G. C., Machado, E. V. A., Constancio, M. T. L., Reis, R. A., & Berchielli, T. T. (2025). Assessing the impact of non-protein nitrogen or rumen undegradable protein supplementation on rumen bacterial diversity and ruminal fermentation in grazing steers during the dry season. Frontiers in microbiology, 16, 1670636. https://doi.org/10.3389/fmicb.2025.1670636

Ezenarro, J., & Schorn-García, D. (2025). How are chemometric models validated? A systematic review of linear regression models for nirs data in food analysis. Journal of Chemometrics, 39(6), e70036. https://doi.org/10.1002/cem.70036

Fan, L., Ma, S., Li, L., & Huang, J. (2024). Fermentation biotechnology applied to wheat bran for the degradation of cell wall fiber and its potential health benefits: A review. International Journal of Biological Macromolecules, 275, 133529. https://doi.org/10.1016/j.ijbiomac.2024.133529

Fodor, M., Matkovits, A., Benes, E. L., & Jókai, Z. (2024). The role of near-infrared spectroscopy in food quality assurance: A review of the past two decades. Foods, 13(21), 3501. https://doi.org/10.3390/foods13213501

Gao, W., Jiang, Q., Guan, Y., Huang, H., Liu, S., Ling, S., & Zhou, L. (2024). Transfer learning improves predictions in lignin content of Chinese fir based on Raman spectra. International Journal of Biological Macromolecules, 269, 132147. https://doi.org/10.1016/j.ijbiomac.2024.132147

García-García, M. D. C., Martín-Expósito, E., Font, I., Martínez-García, B. D. C., Fernández, J. A., Valenzuela, J. L., Gómez, P., & Río-Celestino, M. D. (2022). Determination of quality parameters in mangetout (*Pisum sativum* L. ssp. Arvense) by using vis/near-infrared reflectance spectroscopy. Sensors, 22(11), 4113. https://doi.org/10.3390/s22114113

Gasparini, M., Brambilla, G., Menotta, S., Albrici, G., Avezzù, V., Vitali, R., Buonaiuto, G., Lamanna, M., & Cavallini, D. (2024). Sustainable dairy farming and fipronil risk in circular feeds: Insights from an Italian case study. Food Additives & Contaminants: Part A, 41(12), 1582–1593. https://doi.org/10.1080/19440049.2024.2414954

Gasparini, M., Castrezzati, G., Cavallini, D., Esposito, M., Giugliano, R., Menotta, S., Scortichini, G., Ubaldi, A., Vitali, R., Vivaldi, B., & Brambilla, G. (2026). Fipronil presence in feed materials: Hints for a geographical risk analysis to support the transition towards safe and sustainable food systems. Food Additives & Contaminants: Part A, 43(3), 482–495. https://doi.org/10.1080/19440049.2025.2569060

He, H. J., Zhang, C., Xing, L., Qiao, H., Bi, J., Niu, S., Li, J., Wang, Y., & Ou, X. (2025a). Machine learning-driven pocket-sized NIR analysis for non-invasive quantification of cellulose, hemicellulose, lignin and pectin in sweet potato roots. LWT, 226, 117973. https://doi.org/10.1016/J.LWT.2025.117973

He, H. J., Xing, L., Lv, M., Qiao, H., Wang, Y., Bi, J., Ou, X., & Niu, S. (2025b). Machine learning-assisted micro NIR for rapidly monitoring the variations of enzymatic activity and carbohydrates during saccharification of stored sweet potatoes. LWT, 232(5), 118409. https://doi.org/10.1016/j.lwt.2025.118409

Heidari, F., Øverland, M., Hansen, J. Ø., Mydland, L. T., Urriola, P. E., Chen, C., Shurson, G. C., & Hu, B. (2022). Solid-state fermentation of *Pleurotus ostreatus* to improve the nutritional profile of mechanically-fractionated canola meal. Biochemical Engineering Journal, 187, 108591. https://doi.org/10.1016/j.bej.2022.108591

Jalal, H., Sucu, E., Cavallini, D., Giammarco, M., Akram, M. Z., Karkar, B., Gao, M., Pompei, L., Eduardo, J., Prasinou, P., & Fusaro, I. (2025). Rumen fermentation profile and methane mitigation potential of mango and avocado byproducts as feed ingredients and supplements. Scientific Reports, 15(1), 16164. https://doi.org/10.1038/s41598-025-00675-2

Jiang, H., Mei, C., Li, K., Huang, Y., & Chen, Q. (2018). Monitoring alcohol concentration and residual glucose in solid state fermentation of ethanol using FT-NIR spectroscopy and L1-PLS regression. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 204, 73–80. https://doi.org/10.1016/j.saa.2018.06.017

Jiang, H., Zhang, H., Chen, Q., Mei, C., & Liu, G. (2015a). Identification of solid-state fermentation degree with FT-NIR spectroscopy: Comparison of wavelength variable selection methods of CARS and SCARS. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 149, 1–7. https://doi.org/10.1016/j.saa.2015.04.024

Jiang, H., Zhang, H., Chen, Q., Mei, C., & Liu, G. (2015b). Recent advances in electronic nose techniques for monitoring of fermentation process. World Journal of Microbiology and Biotechnology, 31(12), 1845–1852. https://doi.org/10.1007/s11274-015-1940-0

Jiang, Q., Zhang, M., Mujumdar, A. S., & Wang, D. (2023). Non-destructive quality determination of frozen food using NIR spectroscopy-based machine learning and predictive modelling. Journal of Food Engineering, 343, 111374. https://doi.org/10.1016/j.jfoodeng.2022.111374

Khadem, H., Nemat, H., Elliott, J., & Benaissa, M. (2024). *In vitro* glucose measurement from NIR and MIR spectroscopy: comprehensive benchmark of machine learning and filtering chemometrics. Heliyon, 10(10), e30981. https://doi.org/10.1016/j.heliyon.2024.e30981

Kil, D. Y., Kim, B. G., & Stein, H. H. (2013). Feed energy evaluation for growing pigs. Asian-Australasian Journal of Animal Sciences, 26(9), 1205–1217. https://doi.org/10.5713/ajas.2013.r.02

Li, L., Peng, Y., Li, Y., & Wang, F. (2019). A new scattering correction method of different spectroscopic analysis for assessing complex mixtures. Analytica Chimica Acta, 1087(17), 20–28. https://doi.org/10.1016/j.aca.2019.08.067

Liu, B., Liu, J., Chen, T., Yang, B., Jiang, Y., Wei, D., & Chen, F. (2015). Rapid characterization of fatty acids in oleaginous microalgae by near-infrared spectroscopy. International Journal of Molecular Sciences, 16(4), 7045–7056. https://doi.org/10.3390/ijms16047045

Lozano, M., Rodríguez-Ulibarri, P., Echeverría, J. C., Beruete, M., Sorolla, M., & Beriain, M. J. (2017). Mid-infrared spectroscopy (Mir) for simultaneous determination of fat and protein content in meat of several animal species. Food Analytical Methods, 10(10), 3462–3470. https://doi.org/10.1007/s12161-017-0879-1

Manurung, R., Melinda, R., Abduh, M. Y., Widiana, A., Sugorc, I., & Suheryadi, D. (2015). Potential use of lemongrass (*Cymbopogon winterianus*) residue as dairy cow feed. Pakistan Journal of Nutrition, 14(12), 919–923. https://doi.org/10.3923/pjn.2015.919.923

Massei, A., Cavallini, N., Savorani, F., Falco, N., & Fissore, D., 2025. Exploring NIR spectroscopy data: A practical chemometric tutorial for analyzing freeze-dried pharmaceutical formulations. Chemometrics and Intelligent Laboratory Systems 257, 105291. https://doi.org/10.1016/j.chemolab.2024.105291

Munawar, A. A., Zulfahrizal, & Mörlein, D. (2024). Prediction accuracy of near infrared spectroscopy coupled with adaptive machine learning methods for simultaneous determination of chlorogenic acid and caffeine on intact coffee beans. Case Studies in Chemical and Environmental Engineering, 10, 100913. https://doi.org/10.1016/j.cscee.2024.100913

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology, 46(2), 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024

Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. Analytica Chimica Acta, 1026, 8–36. https://doi.org/10.1016/j.aca.2018.04.004

Ponnampalam, E. N., Jairath, G., Alves, S. P., Gadzama, I. U., Santhiravel, S., Mapiye, C., & Holman, B. W. B. (2025). Sustainable livestock production by utilising forages, supplements, and agricultural by-products: Enhancing productivity, muscle gain, and meat quality – A review. Meat Science, 229(8), 109921. https://doi.org/10.1016/j.meatsci.2025.109921

Rolls, B. J. (2009). The relationship between dietary energy density and energy intake. Physiology & behavior, 97(5), 609–615. https://doi.org/10.1016/j.physbeh.2009.03.011

Saeys, W., Mouazen, A. M., & Ramon, H. (2005). Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. Biosystems Engineering, 91(4), 393–402. https://doi.org/10.1016/j.biosystemseng.2005.05.001

Samadi, S., Kaloudis, E., Wahyudi, I., Wajizah, S., & Munawar, A. A. (2025). Enhancing the accuracy of nirs prediction with adaptive machine learning for assessing fermented citronella feed quality. Tropical Animal Science Journal, 48(5), 440–449. https://doi.org/10.5398/tasj.2025.48.5.440

Samadi, Wahyudi, I., Wajizah, S., Zulfahrizal, & Munawar, A. A. (2023). Rapid and non-destructive prediction of animal feed nutritive parameters using near infrared spectroscopy and multivariate analysis. International Journal of Design and Nature and Ecodynamics, 18(4), 951–956. https://doi.org/10.18280/ijdne.180422

Samadi, Wahyudi, I., Wajizah, S., Zulfahrizal, & Munawar, A. A. (2024). Robust near infrared spectroscopy for rapid and simultaneous determination of fermented cocoa pod husk feed quality attributes. International Journal of Design and Nature and Ecodynamics, 19(2), 379–386. https://doi.org/10.18280/ijdne.190203

Soares, S. F. C., Gomes, A. A., Araujo, M. C. U., Filho, A. R. G., & Galvão, R. K. H. (2013). The successive projections algorithm. TrAC Trends in Analytical Chemistry, 42, 84–98. https://doi.org/10.1016/j.trac.2012.09.006

Sun, Y., Liu, N., Kang, X., Zhao, Y., Cao, R., Ning, J., Ding, H., Sheng, X., & Zhou, D. (2021). Rapid identification of geographical origin of sea cucumbers *Apostichopus japonicus* using FT-NIR coupled with light gradient boosting machine. Food Control, 124, 107883. https://doi.org/10.1016/j.foodcont.2021.107883

Tuyen, D. V., Phuong, H. N., Cone, J. W., Baars, J. J. P., Sonnenberg, A. S. M., & Hendriks, W. H. (2013). Effect of fungal treatments of fibrous agricultural by-products on chemical composition and *in vitro* rumen fermentation and methane production. Bioresource Technology, 129, 256–263. https://doi.org/10.1016/j.biortech.2012.10.128

Wahyudi, I., Munawar, A. A., Kaloudis, E., Wajizah, S., & Samadi. (2025). NIR and machine learning-based rapid monitoring of pH and moisture in citronella residue fermentation. International Journal of Design and Nature and Ecodynamics, 20(11), 2675–2686. https://doi.org/10.18280/ijdne.201119

Wang, Y., Gou, C., Chen, L., Liao, Y., Zhang, H., Luo, L., Ji, J., & Qi, Y. (2023). Solid-state fermentation with white rot fungi (Pleurotus species) improves the chemical composition of highland barley straw as a ruminant feed and enhances *in vitro* rumen digestibility. Journal of Fungi, 9(12), 1156. https://doi.org/10.3390/jof9121156