

Chemometric approach based on feed NIR spectra for rapid assessment of digestibility profiles in unconventional feedstuffs

Indra Wahyudi^{1,2}, *Sitti Wajizah*^{1,2}, *Agus Arip Munawar*^{2,3}, *Efstathios Kaloudis*⁴ and *Samadi Samadi*^{1,2*}

¹Department of Animal Science, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

²Research Center for Innovation and Feed Technology, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

³Department of Agricultural Engineering, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

⁴Computer Simulation, Genomics and Data Analysis Laboratory, Department of Food Science and Nutrition, School of the Environment, University of the Aegean, Myrina 81400, Greece

Abstract. Digestibility profiles serve as vital indicators of feedstuffs' quality, determining how effectively livestock utilise feed nutrients for growth and production. This study develops a rapid and non-destructive model for the digestibility quality of unconventional feedstuffs, which integrates feed NIR spectra with chemometric techniques. A total of 30 samples (citronella residues) obtained from the fermentation process were used, and their NIR spectra (1000–2500 nm) were acquired using a BUCHI NIRFlex N-500 spectrometer. Reference digestibility profiles, including dry matter (IVDMD), organic matter (OM), and pH of rumen fluid, were measured using the two-stage digestion method. Partial least squares regression (PLSR), ridge regression (Ridge), adaptive boosting (AdaBoost), and support vector machine regression (SVMR) models were then constructed to assess digestibility quality. Model performance was assessed using the coefficient of determination (R^2), root-mean-square error (RMSE), and residual predictive deviation (RPD). The results show that the performance of the digestibility profile prediction model meets the accepted standards for NIRS calibration. Notably, the SVMR model exhibits exceptional stability and performance, achieving an R^2 value of ≥ 0.97 and a low RMSE, outperforming the AdaBoost, Ridge, and PLSR models. These findings confirm that the calibration model has significant potential for further independent validation to serve as a rapid-quality assessment tool in analyzing the digestibility profiles of feedstuffs.

* Corresponding author: samadi177@usk.ac.id

1 Introduction

The use of non-conventional fiber-based feedstuffs into ruminant diets presents a strategic solution to the challenges within the livestock feed supply chain. Citronella (*Cymbopogon nardus*) residues are abundant materials, accounting for approximately 99.2% of the total by-products of the essential oil extraction industry [1]. These residues have a substantial fiber content of around 35% [2], indicating their potential as a viable alternative feed source. However, their structurally complex lignocellulosic composition requires biological pretreatment to improve utilization. Fungal fermentation offers a sustainable approach to enhancing digestibility and reducing associated anti-nutritional constraints. Thus, incorporating fermented citronella residues into ruminant diets, not only transforms agro-industrial by-products into a valuable feed resource but also helps reduce environmental impact, lower production costs, and enhance nutritional self-sufficiency by optimizing local resources.

The practical evaluation of feed is of crucial importance to optimise feed efficiency in producers. Accurate assessment of nutritional quality, especially digestibility, is vital for meeting livestock needs. Digestibility is widely regarded as a primary indicator of energy availability in feed [3]. High digestibility is indicative of an effective nutrient provisioning capacity, whereas low digestibility signifies unmet nutritional requirements [4]. Conventionally, digestibility measurements are carried out using *in vivo* and *in vitro* methods, which have become the standard reference. However, these two methods are destructive, time-consuming, costly, and require trained experts [5]. These limitations pose a significant challenge in the process of rapid evaluation and quality control, especially when dealing with large numbers of samples from various fermentation conditions.

Near-infrared spectroscopy (NIRS) has emerged as an efficient and advanced analytical technique due to its ability to perform rapid, non-destructive, and reagent-free analyses. The method relies on the interaction of near-infrared (NIR) light with organic functional groups (e.g., C–H, O–H, and N–H), which produces a characteristic spectrum or "fingerprint" unique to each sample [6]. However, the successful application of NIRS is highly dependent on the accuracy and stability of the calibration model used [7]. Previous studies using the Principal Component Regression (PCR) algorithm have shown the potential of NIRS in feed analysis. However, its current applications are mostly limited to sample classification or ranking, with reported R^2 values of 0.52–0.65 for dry matter digestibility and 0.51–0.66 for organic matter digestibility, indicating that the method is not yet optimized for reliable quantitative prediction [8]. This limitation presents opportunities to develop more accurate calibration models by employing more advanced chemometric using machine learning algorithms.

In this context, the Partial Least Squares Regression (PLSR) algorithm has been the standard method for dealing with multicollinear spectral data [7]. However, recent advancements in machine learning have introduced more adaptable and promising approaches. For instance, Ridge Regression can better handle complex multicollinearity [9], Support Vector Machine Regression (SVMR) is capable of mapping intricate non-linear relationships [10], and Adaptive Boosting (AdaBoost) can enhance prediction accuracy by combining multiple weaker learning models [11]. The four algorithms under consideration encompass both linear (PLSR and Ridge) and non-linear (SVMR and AdaBoost) calibration approaches, thereby facilitating comprehensive model evaluation for a range of data behaviours.

Therefore, this study aims to develop a reliable NIRS calibration model and to evaluate the performance and accuracy of four machine learning algorithms in creating a prediction model for the quality of fermented feed based on citronella residues.

2 Material and Method

2.1 Description of the feedstuff sample

This study utilized processed feed derived from unconventional materials, specifically citronella residues. A total of 30 samples were collected at the final stage of a sequential fermentation process, in which fermenting material from one batch served as the inoculum for the subsequent batch. Briefly, fermentation was initiated by homogenizing 447 g of dried citronella residue with distilled water to achieve a 60% moisture content, followed by the addition of 30 g of molasses and 100 g of corn bran. The substrate matrix was inoculated with various fungal strains, including *Phanerochaete chrysosporium*, *Pleurotus ostreatus*, *Trichoderma viride*, and *Lentinula edodes*, and fermented under aerobic conditions at room temperature ($\approx 37^\circ\text{C}$) for 28 days. Subsequently, the substrate was ensiled with *Lactobacillus buchneri* under anaerobic conditions for 14 days. All treatments, including the control, were conducted in six independent replicates using a completely randomized design. Following fermentation, all substrates were oven-dried at 60°C to obtain dry material for chemical analysis and NIR spectral measurements.

2.2 NIR spectra acquisition

The NIR spectra of the processed feed samples were obtained using the Büchi NIRFlex N-500 spectrometer (Büchi Italia, Cornaredo, Italy). Approximately 30 g of finely ground feed samples were used for spectral analysis. The NIR spectra was recorded in reflection mode by averaging 32 scans over a wavelength range of 1,000 to 2,500 nm, with intervals of 2 nm [12]. All measurements were conducted at room temperature.

2.3 Measurement of digestibility profiles

Subsequent to the collection of all spectral data, the same samples were analyzed following laboratory procedures to assess in vitro digestibility. This included measuring dry matter digestibility (IVDMD), organic matter digestibility (IVOMD), and the pH of rumen fluid. The in vitro digestibility was determined using the Tilley and Terry two-stage digestion method [13]. Following the completion of in vitro fermentation, the pH of the rumen fluid was measured using a pH meter. Each sample was analyzed in duplicate to ensure analytical repeatability.

2.4 Chemometrics and models evaluations

Chemometric modelling was employed using the full spectral range of raw data to establish predictive digestibility profiles. Four machine learning algorithms were employed: Partial Least Squares Regression (PLSR), Ridge Regression (Ridge), Adaptive Boosting (AdaBoost), and Support Vector Machine Regression (SVMR). It is evident that each method possesses a distinct set of advantages, allowing for the handling of different types of data features and model complexities. These advancements encompass both linear (PLSR and Ridge) and non-linear (SVMR and AdaBoost) calibration approaches, facilitating a more comprehensive model evaluation. In this study, external data partitioning was not applied due to the limited sample size. Instead, model performance was tested using internal cross-validation, where all samples were used in rotation for both calibration and validation. This method helps estimate how well the model generalizes without reducing the number of samples available for calibration. Model performance was evaluated based on standard calibration and validation metrics, including the coefficient of correlation (r), the coefficient

of determination (R^2), the Root Mean Square Error (RMSE), the Residual Predictive Deviation (RPD) and the Range Error Ratio (RER) to assess accuracy, robustness, and predictive capability [12].

3 Results and Discussion

3.1 Digestibility profiles of the sample dataset

A total of 30 sequentially fermented feed samples were analyzed, representing five treatments: four fungal inoculations (*P. chrysosporium*, *P. ostreatus*, *T. viride*, and *L. edodes*) and a non-inoculated control. All samples underwent 28 days of aerobic fungal fermentation followed by 14 days of anaerobic ensiling with *L. buchneri*. The descriptive statistics for the processed samples are summarized in Table 1, covering in vitro dry matter digestibility (IVDMD), in vitro organic matter digestibility (IVOMD), and pH values obtained using the standard two-stage digestion method.

Table 1. Descriptive statistics of actual measurements of feed digestibility in the data set (n=30)

Feed digestibility	Range	Minimum	Maximum	Mean	SD	Variance
IVDMD (%)	6.14	36.35	42.49	39.70	1.62	2.62
IVOMD (%)	7.16	33.90	41.06	38.01	1.90	3.63
pH	0.13	6.85	6.98	6.92	0.04	0.00

Note: IVDMD, in vitro dry matter digestibility; IVOMD, in vitro organic matter digestibility; SD, standard deviation.

As shown in Table 1, moderate variation was observed in the IVDMD and IVOMD parameters, with average values of 39.70% and 38.01%, respectively, and ranges of 6.14 and 7.16. For feed analysis, a digestibility range of 5 to 10% points aligns with the general recommendations for differences between samples in feed analysis. This range not only helps differentiate the variability in spectral responses, which is crucial for developing a reliable NIRS calibration model. Sufficient variability in these parameters is vital because the effectiveness of NIRS calibration is highly dependent on the diversity of analyte concentrations in the calibration data set [14]. A varied range of values allows the model to recognize and learn spectral characteristics associated with differences in feed digestibility, thereby improving the accuracy and reliability of predictions [4].

Conversely, the minimal variability observed in pH values, with a range of 0.13 and a variance close to 0.00, suggests that pH may offer limited spectral differentiation for NIRS calibration in this specific context. NIRS works by correlating near-infrared light absorption with the chemical composition and physical properties of samples [6]. In circumstances where parameters such as pH demonstrate minimal variation, it becomes challenging to distinguish their distinct spectral signatures and establish a correlation with NIRS data. This limitation has the potential to diminish their contribution to the overall predictive model. Therefore, while IVDMD and IVOMD are essential nutritional parameters for evaluating feed, the narrow range of pH suggests that it may not be a significant factor in the spectral differences observed in this dataset.

3.2 Prediction of digestibility profiles

The IVDMD and IVOMD are widely accepted as essential indicators for evaluating the nutritional quality of grass and feed materials for ruminants. These parameters are typically measured using the two-stage in vitro fermentation method of Tilley and Terry [13], which simulates rumen and post-rumen digestion to estimate digestible dry and organic matter.

Higher lignin-to-cellulose ratios are negatively correlated with IVDMD and IVOMD. Therefore, this consistent relationship between feed chemical characteristics and digestibility further emphasizes the relevance of IVDMD and IVOMD as reference parameters in the NIRS calibration process for rapid and non-destructive feed quality evaluation. The prediction of IVOMD and IVDMD by NIRS is indirect, relying on the interaction of infrared light with organic molecular bonds such as C–H, O–H, and N–H present in the feed matrix. The performance of the NIR predictive model for digestibility profiles using four machine learning algorithms is summarized in Table 2 and visualized in Fig. 1.

Table 2. Prediction performance of PLSR, Ridge, AdaBoost, and SVMR models in determining digestibility of fermented citronella residues

Feed digestibility	Method	r	R ²	RMSE	RPD	RER
IVDMD	PLSR	0.92	0.85	0.61	2.66	10.07
	Ridge	0.92	0.85	0.61	2.66	10.07
	AdaBoost	0.99	0.99	0.16	10.12	38.38
	SVMR	0.99	0.99	0.15	10.80	40.93
IVOMD	PLSR	0.91	0.84	0.75	2.54	9.55
	Ridge	0.91	0.84	0.75	2.54	9.55
	AdaBoost	0.99	0.99	0.15	12.70	47.73
	SVMR	0.99	0.99	0.14	13.60	51.14
pH	PLSR	0.95	0.91	0.01	4.39	13.00
	Ridge	0.97	0.96	0.01	5.49	16.25
	AdaBoost	0.98	0.97	0.01	6.27	18.57
	SVMR	0.98	0.97	0.01	6.27	18.57

Note: IVDMD, in vitro dry matter digestibility; IVOMD, in vitro organic matter digestibility, PLSR, partial least squares regression; Ridge, ridge regression; AdaBoost, adaptive boosting; SVMR, support vector machine regression; r, coefficient of correlation; R², coefficient of determination; RMSE, root mean square error; and RPD, residual predictive deviation; RER, range error ratio.

As demonstrated in Table 2, the performance of the digestibility profile prediction model meets the accepted standards for NIRS calibration. Based on literature, RPD values greater than 2, the NIRS calibration model demonstrates sufficient predictive stability and can make reliable forecasts. The results clearly show that ensemble-based and kernel-based models (SVMR and AdaBoost) consistently outperform linear models (PLSR and Ridge) on all digestibility profile parameters (Fig. 1). For IVDMD prediction, SVMR achieved the highest accuracy, with with a high R² (0.99), a very low RMSE (0.15), and a high RPD value (10.80), indicating good calibration model reliability. Similarly, for IVOMD, SVMR also showed optimal performance (R²= 0.99, RMSE= 0.14, RPD= 13.60), followed by AdaBoost (R²= 0.99, RMSE= 0.15, RPD= 12.70). In the context of pH values, SVMR once again demonstrated its superior predictive capability (R²= 0.97, RMSE= 0.01). The pH model shows a higher correlation value than expected given the narrow variation in the dataset (Table 1). However, it should be noted that this apparent performance should be interpreted with caution. The enhancement of performance metrics, such as high R² and RPD values, on the basis of limited variation in reference values, results in an overestimation of the model's reliability. In such cases, the model may appear accurate not as a result of its ability to capture meaningful spectral-chemical relationships but rather due to the narrow reference range, which enables the model to accommodate noise or random fluctuations. This phenomenon can be interpreted as an indication of overfitting, a common challenge encountered in NIRS modeling when analyte variability is minimal. In order to enhance the performance of future models, it is imperative to consider a broader range of pH values and the validation of independent batches.

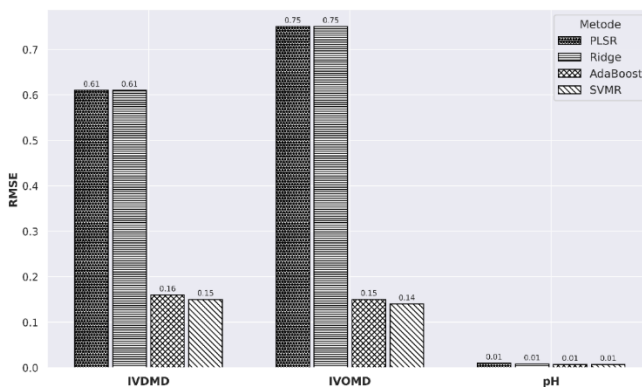


Fig 1. Performance comparison among four machine learning algorithms in term of RMSE for digestibility profile predictions.

Linear models such as PLSR and Ridge perform effectively by leveraging multivariate information across the entire NIR wavelength range and employing regularization techniques to mitigate the risk of overfitting. However, these linear approaches may fall short in scenarios where the relationship between spectral data and feed parameters is non-linear and subtle. In contrast, non-linear models like SVMR and AdaBoost excel at capturing these intricate patterns. The SVMR model utilizes kernel functions to transform spectral data into a higher-dimensional space [10], allowing even small variations to become more discernible. Similarly, AdaBoost enhances the detection of weak signals by combining multiple simple models [11], thereby improving sensitivity to minor variations, as indicated by elevated R² and RPD values for IVDMD and IVOMD. However, in circumstances where the available training data is severely restricted, such as in the case of pH values, ML models tend not only to discern relevant patterns but also to capture noise and random fluctuations. This phenomenon increases the risk of overfitting. Fig 2. shows the scatter plots of predicted versus actual values for digestibility profiles using optimal machine learning models (Table 2 and Fig. 1).

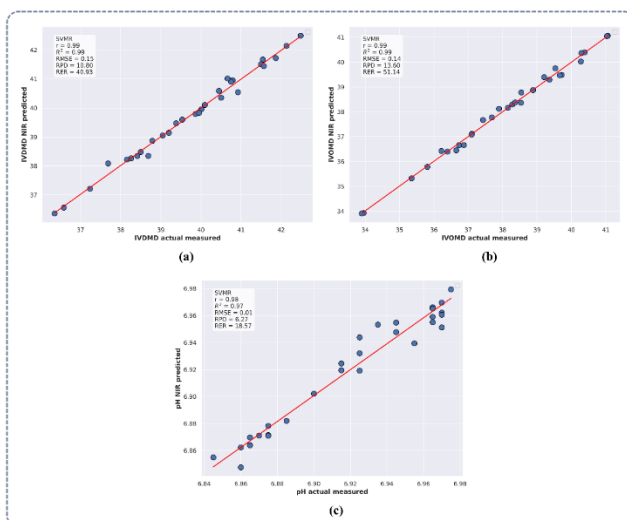


Fig. 2. Scatter plots of predicted versus actual digestibility values using optimal machine learning models: (a) IVDMD by SVMR, (b) IVOMD by SVMR, and (c) pH by SVMR. Data points show the agreement between predictions and reference values, with the red line ($y = x$) indicating perfect prediction.

4 Conclusion

This study demonstrates that integrating NIRS with machine learning-based chemometric approaches offers a promising and rapid method for assessing feed digestibility. The calibration models demonstrated performance that met the criteria for NIRS prediction. Among the evaluated models, SVMR demonstrated the highest predictive accuracy, followed by AdaBoost, both achieving R^2 values above 0.97 and high RPD scores, thereby emphasising their robustness. Nonetheless, when the variability of training data is significantly restricted, as seen with the pH parameter, ML models may capture noise in addition to relevant patterns, thereby increasing the risk of overfitting. To enhance model performance moving forward, it is essential to incorporate a wider range of data and implement independent batch validation to ensure the model's applicability.

Acknowledgements

This research was supported by Universitas Syiah Kuala for the fiscal year 2024.

Fundings

This research was supported by Universitas Syiah Kuala for the fiscal year 2024.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contribution statement

Indra Wahyudi: Conceptualization, Methodology, Investigation, Writing, Original draft preparation. Sitti Wajizah: Data curation, Writing- Original draft preparation. Agus Arip Munawar: Validation, Writing, Reviewing and Editing. Efstathios Kaloudis: Validation, Writing, Reviewing and Editing. Samadi Samadi: Conceptualization, Methodology, Investigation, Validation, Writing, Reviewing and Editing.

References

1. R. Manurung, R. Melinda, M.Y. Abduh, A. Widiانا, I. Sugorc, D. Suheryadi, Potential use of lemongrass (*Cymbopogon winterianus*) residue as dairy cow feed. *Pak. J. Nutr.* **14**, 919–923. (2015). <https://doi.org/10.3923/PJN.2015.919.923>
2. Sari, W. Manguwardoyo, I. Sugoro, Degradasi ampas dan serai wangi segar (*Cymbopogon nardus* L) dengan metode in sacco pada kerbau fistula. *Prosiding Seminar Nasional Teknologi Peternakan dan Veteriner*, 118–124 (2017). <https://doi.org/10.14334/pros.semnas.tpv-2017-p.119-125>
3. D. Susanti, N. Jamarun, F. Agustin, T. Astuti, G. Yanti, Kecernaan in-vitro fraksi serat kombinasi pucuk tebu dan titonia fermentasi sebagai pakan ruminansia. *J. Agripet.* **20**, 86–95 (2020). <https://doi.org/10.17969/agripet.v20i1.16040>
4. N.F. Sari, R. Ridwan, R. Rohmatussolihat, R. Fidriyanto, W.D. Astuti, Y. Widyastuti, Characteristic of different level of fermented concentrate in the rumen metabolism based on in vitro. *J. Indones. Trop. Anim. Agric.* **43**, 296–305 (2018). <https://doi.org/10.14710/jitaa.43.3.296-305>
5. R. Andrés-Camelo, J. Mazabel, P. Espitia-Buitrago, R.N. Jauregui, J.A. Cardoso, Near-infrared spectroscopy and wet chemistry dataset for forage nutritional quality assessment in *Urochloa humidicola*. *Data Brief.* **60**, 111651 (2025). <https://doi.org/10.1016/j.dib.2025.111651>
6. D. Cozzolino, The ability of near infrared (NIR) spectroscopy to predict functional properties in foods: Challenges and opportunities. *Molecules.* **26**, 6981 (2021). <https://doi.org/10.3390/molecules26226981>
7. C. Maduro Dias, H. Nunes, A. Borba, Near-infrared spectroscopy in animal nutrition: Historical insights, technical principles, and practical applications. *Analytica.* **5**, 481–498 (2024). <https://doi.org/10.3390/analytica5040033>

8. Samadi, E.S. Pratama, S. Wajizah, Ilham, I. Wahyudi, A.A. Munawar, The use of near-infrared spectroscopy to rapidly determine in vitro digestibility profiles in fermented agro-waste ruminant diets. *IOP Conf. Ser. Earth Environ. Sci.* **1477**, 012001 (2025). <https://doi.org/10.1088/1755-1315/1477/1/012001>
9. M.F. Alharthi, N. Akhtar, Modified two-parameter ridge estimators for enhanced regression performance in the presence of multicollinearity: Simulations and medical data applications. *Axioms*. **14**, 527 (2025). <https://doi.org/10.3390/axioms14070527>
10. R. Rodríguez-Pérez, J. Bajorath, Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *J. Comput. Aided Mol. Des.* **36**, 355–371 (2022). <https://doi.org/10.1007/s10822-022-00442-9>
11. S.S. Hussain, S.S.H. Zaidi, AdaBoost ensemble approach with weak classifiers for gear fault diagnosis and prognosis in DC motors. *Appl. Sci.* **14**, 3105 (2024). <https://doi.org/10.3390/app14073105>
12. Samadi, I. Wahyudi, S. Wajizah, Zulfahrizal, A.A. Munawar, Robust near infrared spectroscopy for rapid and simultaneous determination of fermented cocoa pod husk feed quality attributes. *Int. J. Des. Nat. Ecodyn.* **19**, 379–386 (2024). <https://doi.org/10.18280/ij dne.190215>
13. J.M.A. Tilley, R.A. Terry, A two-stage technique for the in vitro digestion of forage crops. *Grass Forage Sci.* **18**, 104–111 (1963). <https://doi.org/10.1111/j.1365-2494.1963.tb00335.x>
14. X. Yang, A. Arroyo Cerezo, P. Berzaghi, L. Magrin, Comparative near infrared (NIR) spectroscopy calibrations performance of dried and undried forage on dry and wet matter bases. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **316**, 124287 (2024). <https://doi.org/10.1016/j.saa.2024.124287>