International Journal of Design & Nature and Ecodynamics
Vol. 21, No. 7, July, 2026, pp. 1957-1969
Journal homepage: http://iieta.org/journals/ijdne

Machine Learning-Enhanced NIR Spectroscopy for Intelligent Monitoring of Lignocellulosic Feedstuffs Biodegradation by Lignocellulolytic Fungi

Indra Wahyudi[1,2,3], Agus Arip Munawar[3,4], Efstathios Kaloudis[5], Anjas Asmara Samsudin[6,7], Sitti Wajizah[2,3], Samadi[2,3*]

[1] Doctoral Program of Agricultural Science, Postgraduate School, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[2] Department of Animal Science, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[3] Research Center for Innovation and Feed Technology, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[4] Department of Agricultural Engineering, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[5] Computer Simulation, Genomics and Data Analysis Laboratory, Department of Food Science and Nutrition, School of the Environment, University of the Aegean, Myrina 81400, Greece
[6] Department of Animal Science, Faculty of Agriculture, Universiti Putra Malaysia, Serdang 43400, Malaysia
[7] Institute of Tropical Agriculture and Food Security, Universiti Putra Malaysia, Serdang 43400, Malaysia

Corresponding Author Email: samadi177@usk.ac.id

https://doi.org/10.18280/ijdne.210710

Received: 15 May 2026
Revised: 12 July 2026
Accepted: 25 July 2026
Available online: 31 July 2026

Keywords:
near-infrared, machine learning, lignocellulosic biomass, biodegradation, monitoring, white-rot fungi

ABSTRACT

This study developed a machine learning (ML)-enhanced near-infrared spectroscopy (NIRS) framework for monitoring lignocellulosic biomass (LCB) biodegradation during solid-state fermentation (SSF) by white-rot fungi. Distinctive species, including *P. chrysosporium*, *P. ostreatus*, *T. viride*, and *L. edodes*, were used in the 28-day biodegradation of citronella residues (CR). A Successive Projections Algorithm (SPA) was used to select informative wavelengths (12–21 wavelengths), and five ML models were evaluated. Random Forest (RF) achieved the strongest predictive performance for moisture (R^2_V = 0.80; $RMSE_V$ = 1.02%), pH (R^2_V = 0.73; $RMSE_V$ = 0.63), gross energy (R^2_V = 0.78; $RMSE_V$ = 0.13 MJ/kg DM), protein (R2V = 0.73; $RMSE_V$ = 0.32%), and acid detergent lignin (ADL) (R^2_V = 0.72; $RMSE_V$ = 0.48%). In contrast, all models failed to predict neutral detergent fiber (NDF), acid detergent fiber (ADF), and cellulose (negative R^2_V), highlighting the fundamental limitations related to the spectral overlap and structural similarity of polysaccharide fractions. These findings demonstrate the feasibility of NIRS-ML for selective process monitoring while identifying key constraints in predicting polysaccharide fractions. This study provides a basis for developing NIR-based process analytical tools for SSF systems.

1. INTRODUCTION

The use of lignocellulosic biomass (LCB) derived from biodegradation as a feed source for ruminants has emerged as a sustainable resource that helps reduce waste and address environmental challenges. LCB, comprising agro-industrial waste such as citronella residues (CR), is widely available in tropical regions. Notably, these residues constitute a dominant proportion of waste from the essential oil extraction industry, accounting for approximately 99.2% of total by-products [1]. At its core, this biomass is composed of cellulose (40–60%), hemicellulose (20–30%), and lignin (10–25%), which are bound together in a complex matrix [2]. These polymers are commonly evaluated using detergent fiber fractionation: neutral detergent fiber (NDF), acid detergent fiber (ADF), and acid detergent lignin (ADL), which reflect feed quality and digestibility [3]. Biodegradation is a strategic approach to overcoming the recalcitrance of plant cell wall structures, which severely limits their nutritive value for livestock.

From a livestock nutrition perspective, lignin acts as an antinutritional factor. Lignin forms ester and ether bonds with hemicellulose and cellulose, thereby blocking access to polysaccharides by rumen microbial enzymes [3]. Furthermore, its degradation produces phenolic compounds, such as p-coumaric and ferulic acids, which inhibit the activity of cellulolytic microbes [4]. Consequently, the digestibility of NDF and ADF decreases, feed intake is reduced, and energy and protein utilization in ruminants becomes suboptimal. Therefore, the reduction of ADL content through processing is the primary biological objective of lignocellulosic feedstuffs biodegradation. Nowadays, biodegradation mediated by white-rot fungi via solid-state fermentation (SSF) has emerged as a highly promising approach for addressing these challenges, as it is an eco-friendly way to convert LCB into animal feed [5]. White-rot fungi generate a range of ligninolytic enzymes, including laccase, lignin peroxidase (LiP), and manganese peroxidase (MnP), which selectively depolymerize the lignin through oxidative free-radical mechanisms [6, 7]. This process releases the bound polysaccharide fractions, making them more accessible and

significantly improving the nutritional value and rumen fermentability of LCB.

To maximize the nutritional value of LCB through white-rot fungal biodegradation, a temporal monitoring framework is essential for tracking compositional changes throughout the bioconversion process. The SSF period is a critical phase during which fungal mycelium gradually colonizes the substrate [8, 9]. This process involves the release of ligninolytic and cellulolytic enzymes that modify the NDF, ADF, and ADL fractions, depending on both the incubation stages and the fungal strains used [10]. Systematic monitoring is a prerequisite for the future identification of the treatment duration that maximises lignin modification while protecting cellulose and hemicellulose from consumption. Meanwhile, cellulose and hemicellulose (the primary energy source for rumen microbiota) must be protected from consumption, a phenomenon well documented in cases of non-selective fungal degradation [11]. Beyond fiber fractions, simultaneous monitoring of protein is equally indispensable, as SSF simultaneously promotes protein enrichment through fungal biomass accumulation [12]. Therefore, sequential sampling and non-destructive monitoring are critical for tracking these compositional changes over time [10, 13].

Several standard analytical methods, such as the Van Soest method and proximate analysis, have been used to determine fiber fractions and nutritional components [14]. Although these methods are well-established for evaluating feed quality, they are fundamentally unsuited to continuous, dynamic biological process monitoring. This procedure requires physical samples and involves chemical digestion [15], which affects the structural integrity of the fermentation substrate. Near-infrared spectroscopy (NIRS) has emerged as a compelling solution to this challenge. NIRS reveals chemical information based on the distinct absorption and scattering of light in the near-infrared (NIR) region (750-2500 nm) by various chemical compounds such as C–H, O–H, N–H, and S–H [16-18]. Biodegradation by white-rot fungi during SSF alters the presence and intensity of specific functional groups, thereby altering NIR spectra [19]. These characteristics can be used to detect and distinguish various fermentation stages. Compared to standard methods, NIRS is a fast, non-destructive technique, particularly important for real-time monitoring of the LCB biodegradation process.

To the best of our knowledge, few studies have applied NIRS-ML frameworks to monitor the biodegradation of LCB by white rot fungi during SSF. Meanwhile, NIR spectroscopy has been progressively adopted for fermentation monitoring in food-related systems. For instance, FT-NIR spectroscopy has been successfully employed to monitor the growth of *P. ostreatus* and *G. annularis* by detecting compositional shifts in food by-products in SSF [20]. Therefore, this study aimed to develop and validate NIR models for the intelligent monitoring of fiber fraction dynamics and key nutritional compositional changes in citronella LCB throughout an SSF process mediated by white rot fungi. Specifically, this study aimed to: (1) characterise the compositional dynamics (fiber fractions and key nutritional parameters) of CR across a 28-day SSF by four lignocellulolytic fungi; (2) employ the Successive Projections Algorithm (SPA) for wavelength selection to enhance model simplicity, diminish spectral dimensionality, and enhance physicochemical interpretability; and (3) systematically develop, assess, and compare the predictive efficacy of five chemometric and advanced machine learning (ML) algorithms, namely Partial Least Squares Regression (PLSR), Random Forest (RF), Support Vector Regression (SVR), Adaptive Boosting (AdaBoost), and Gradient Boosting Regression (GBR), while identifying the optimal modelling framework for each target analyte.

2. MATERIALS AND METHODS

2.1 Substrate and spawn preparation

CR from the oil extraction industry was obtained from farmers (Gayo Lues District of Aceh, Indonesia) and chopped to an average particle length of approximately 3 cm. The chopped material was dried in a forced-air oven at 60 ℃ for approximately 48 h until the moisture content reached 10–12% (w/w), preserving the structural carbohydrates without thermal degradation [21]. Four lignocellulolytic fungi, *Phanerochaete chrysosporium* (InaCC F206; strain A), *Pleurotus ostreatus* (InaCC F110; strain B), *Trichoderma viride* (InaCC F241; strain C), and *Lentinula edodes* (InaCC F95; strain D) were used in this study. All fungal cultures were procured from the Indonesian Culture Collection (InaCC) Laboratory, Badan Riset dan Inovasi Nasional (BRIN), Indonesia. Prior to fermentation, each strain was activated by culturing on Potato Dextrose Agar (PDA) plates at 24 ℃ until mycelia completely colonized the medium, as described by research [22]. Then, spawns were prepared by placing cultured agar (1 cm × 1 cm) into sterilized corn grain and incubating at 24 ℃ until mycelia colonized all grains. Completed spawns were stored in a refrigerator (6 ℃) until required for the fermentation experiment.

2.2 Experimental design and solid-state fermentation procedure

The experiment was a completely randomized design, comprising four fungal treatments and an uninoculated control (no strain) in three biological replicates. SSF was performed at pilot scale using sealed polyethylene bags equipped with several 2-mm-diameter ventilation holes to facilitate aerobic conditions. Gas exchange occurred by passive diffusion through the perforations, with no active aeration or bag agitation during fermentation. Each bag contained 447 g of CR, 30 g of molasses, and 100 g of corn bran. All components were thoroughly mixed to achieve homogeneity. Subsequently, the substrate was inoculated with 50 g of spawn for each fungus (strains A, B, C, and D). Sterile water was added during mixing to maintain an overall moisture level of 60% (w/w). The uninoculated control received the same basal substrate and was adjusted to the same moisture content, but did not receive spawn or any spawn-equivalent addition. It was formulated as the unfermented substrate base, allowing comparative data against which changes associated with fungal activity in the inoculated treatments could be evaluated. Both the SSF and the uninoculated substrate (no strain) were incubated at ambient conditions, without active temperature control. The temperature averaged 37 ℃, fluctuating naturally (±3 ℃) during the 28-day fermentation period. Continuous ambient and substrate-temperature records and relative-humidity logs were not maintained during the fermentation period.

2.3 Fermentation periods and sampling strategy

The 28-day SSF was partitioned into four consecutive 7-day sampling periods (days 7, 14, 21, and 28) to monitor temporal changes in nutritional composition throughout the SSF process. Five treatment conditions (four fungal strains and one uninoculated control) were prepared, with three independent biological replicate fermentation bags per treatment for each sampling period. To avoid disturbing the fermentation process, a sacrificial sampling design was used, in which each fermentation bag was opened only once at its designated sampling time. Consequently, a total of 60 independent fermentation bags were prepared for the experiment (4 sampling periods × 5 treatments × 3 biological replicates = 60 bags). At each sampling period, the corresponding 15 bags (5 treatments × 3 replicates) were aseptically opened, and a representative subsample was collected from each bag for NIR spectral acquisition and subsequent reference chemical analysis, yielding 60 independent samples across the entire experiment. This sampling strategy ensured that each observation represented an independent biological replicate without repeated sampling of the same fermentation bag, thereby providing an appropriate dataset for evaluating temporal changes in biomass composition and for subsequent ML analyses. Although suitable for an initial proof-of-feasibility study, the dataset remains relatively limited for establishing broad model generalizability.

2.4 Near-infrared spectra acquisition

NIR spectral measurements were performed immediately following the aseptic collection of each sample at each time point. Approximately 5 g of each SSF sample was placed in the sample holder as a single presentation and flattened to create a smooth surface for scanning; the holder was rotated, and a single subsample was analysed per fermentation bag. Spectra were acquired using a NIRFlex N-500 spectrometer (Büchi, Flawil, Switzerland) operating in reflectance mode, with absorbance (log 1/R) recorded over the 1000–2500 nm range. As a Fourier-transform instrument, the NIRFlex N-500 produces spectra sampled on a grid that is equally spaced in wavenumber (a constant step of ≈ 3.86 cm^{-1}), which corresponds to a non-uniform spacing in wavelength that increases from ≈ 0.4 nm near 1000 nm to ≈ 2.4 nm near 2500 nm. Each spectrum comprised 1,557 spectral channels, which were used directly as input to wavelength selection without resampling onto a uniform wavelength grid. Wavelength positions are reported in nm (rounded to the nearest nm) for readability. Each reported spectrum was obtained by co-averaging 32 consecutive scans to improve the signal-to-noise ratio; these co-added scans therefore characterise instrumental repeatability rather than within-sample (sampling) repeatability. All measurements were performed at an ambient temperature of approximately 29–31 °C.

2.5 Chemical analyses

Following NIR spectral acquisition, each sample was subjected to a comprehensive suite of reference chemical analyses. Chemical analyses were used to determine the moisture content, pH, gross energy, protein, and fiber fractions, including NDF, ADF, ADL, and cellulose. The moisture content was determined gravimetrically by oven-drying at 105 °C until a constant mass was achieved. Substrate pH was measured by preparing a 1:10 (w/v) suspension of the sample in distilled water and determining it using a calibrated pH meter. Gross energy was determined using adiabatic Bomb Calorimetry. Protein (N × 6.25) was quantified using the Kjeldahl method according to AOAC International procedures [23]. Fiber fractions, including NDF, ADF, and ADL, were determined by the sequential Van Soest detergent extraction method [14]. Cellulose content was subsequently calculated as the difference between ADF and ADL values.

2.6 Dataset partitioning and spectral enhancement

The datasets were split into calibration and validation sets with a ratio of 3:1, comprising 45 samples used to construct calibration models and 15 samples used to validate the calibration models. The dataset was randomly stratified to obtain a proportional representation of the reference value distributions across both sets. All NIR spectra were pre-processed using Multiplicative Scatter Correction (MSC). MSC was selected because SSF substrates are inherently heterogeneous in terms of particle size, surface texture, and bulk density. These physical properties introduce both additive and multiplicative scattering effects. The MSC is widely used and effectively eliminates spectral variations caused by sample scattering, thereby enhancing the correlation between spectra and reference data [24].

It should be acknowledged that the total sample size (n = 60, comprising 45 calibration and 15 validation samples) is limited for developing and tuning five ML models, particularly the ensemble-based AdaBoost and GBR algorithms, for which 45 calibration samples create a substantial risk of overfitting. Because of this constraint, model performance in the present study is reported from a single stratified calibration/validation split (Section 3.5) rather than from repeated k-fold cross-validation, which would provide mean ± standard deviation (SD) estimates of predictive performance across resampled folds and a more robust characterisation of variance than a single split can offer. Repeated cross-validation was not implemented here because, with only 45 calibration samples, partitioning into further folds would leave individual folds too small to yield stable estimates for the ensemble methods. Consequently, the calibration and validation coefficients of determination (R^2_C and R^2_V) values reported throughout this study should be interpreted as preliminary estimates of predictive capacity rather than as robust, generalisable measures of model performance, and repeated k-fold cross-validation on an expanded dataset is identified as a priority for future work.

2.7 Wavelength selection via Successive Projections Algorithm

The most informative wavelengths from the NIR spectra were selected using the SPA. The SPA is a forward variable selection method that iteratively selects wavelength subsets. This approach reduces multicollinearity among selected variables while retaining the most relevant variables for predictive modelling. SPA is particularly suitable for high-dimensional datasets, such as NIR spectra, in which the number of wavelengths greatly exceeds the number of samples [25]. SPA was applied independently to each of the parameters using the calibration dataset to generate candidate wavelength subsets through QR-based projection. For each potential number of selected wavelengths (from 1 up to a user-defined

maximum of 40), the SPA performed a forward selection process and was evaluated using prediction residual error sum of squares (PRESS). For each candidate subset size (1 up to 40), performance was assessed by cross-validation within the calibration set only, and the optimal number of wavelengths was selected as that minimising the cross-validated RMSE (equivalently, PRESS). The independent validation set was not used at any stage of wavelength selection. This calibration-only procedure was adopted to reduce overfitting and to keep the held-out validation set fully independent of model development.

2.8 Machine learning modelling

The SPA-selected wavelengths were used as predictors, whereas the moisture content, pH, gross energy, protein, and fiber fractions (NDF, ADF, ADL, and cellulose) were used as response variables. In this study, five ML methods were employed and evaluated: PLSR, RF, SVR, AdaBoost, and GBR. PLSR was included as the established chemometric baseline, constructed by extracting latent variables (LVs) that maximize the covariance between the spectral (X) and reference (Y) matrices [26, 27]. An SVR with a radial basis function (RBF) kernel was employed to capture potential nonlinear relationships that may not be adequately represented by the LV decomposition of PLSR [1, 28]. The SVR model was fine-tuned by optimizing the regularization parameter (C), kernel coefficient (γ), and epsilon-insensitive loss parameter (ε). RF is a bagged ensemble algorithm that improves predictive performance by constructing a set of decision trees, each trained on randomly sampled subsets of the original dataset via bootstrap sampling [29]. The RF was tuned by optimizing *n_trees*, while its mean decrease in node impurity served as a post-hoc validation of the SPA-selected wavelengths. Adaboost is a sequential ensemble method that fundamentally involves iterative weight adjustments for samples, particularly focusing on misclassified samples in previous rounds [30]. Adaboost was optimized by tuning the *learning_rate* and *n_estimators* to control the learner contribution and error correction. The GBR extends the boosting framework by fitting each subsequent estimator to the negative gradient of the loss function of the current ensemble, enabling flexible, high-accuracy regression through the sequential correction of residual prediction errors [31]. The GBR was tuned with *n_estimators*, *learning_rate*, *max_depth*, and the L2 regularization parameter (λ) was optimized to balance model complexity and generalization. All hyperparameter optimisation was performed by cross-validation restricted to the calibration set (n = 45). The validation set (n = 15) was reserved exclusively for the single final evaluation and was not used for feature selection or hyperparameter tuning.

2.9 Model evaluation

Model development (wavelength selection and hyperparameter tuning) was confined to the calibration set, and the validation set was used only once for final assessment; this strict separation prevents information from the validation samples from influencing model construction. The model performance was evaluated using the R^2_C and R^2_V and the corresponding root mean square errors ($RMSE_C$ and $RMSE_V$). An R^2 value greater than 0.80 is indicative of strong predictive capacity, whereas an RMSE lower than the reference SD reflects satisfactory accuracy [17, 32]. In NIRS applications, there is no universal threshold for "acceptable" or "good" prediction performance, as criteria depend on the analyte, matrix complexity, and intended use. However, the following conventions drawn from chemometrics and food/feed NIRS studies were adopted here as reference guidelines: $R^2 \geq 0.90$ is generally considered excellent and suitable for routine quality assurance [33]. $0.70 \leq R^2 < 0.90$ indicates a useful prediction for screening purposes or process monitoring, provided the RMSE is within practical limits [34]. $R^2 < 0.50$ suggests poor predictive ability, and the model may not be reliable even for ranking samples [35, 36].

Pearson correlation analysis was used to examine associations among substrate parameters and fermentation time (Section 3.2) as a descriptive, exploratory tool. No formal treatment × time statistical analysis (e.g., mixed-model or factorial Analysis of Variance (ANOVA)) was performed; consequently, the reported correlations indicate association rather than causation and cannot be used to draw conclusions about strain-specific degradation differences.

2.10 Computational and visualization tools

Spectral preprocessing and wavelength selection were performed in Python 3.13 within the Google Colab environment (12 GB RAM). Data handling and matrix operations were performed using the *Pandas* and *NumPy* libraries. The SPA algorithm was implemented using the *SciPy* module. The pre-processed data was then imported into Orange Data Mining (version 3.40) software to build ML models. The optimized models were exported and reloaded into the Google Colab environment using the *orange3* library. All graphical outputs were generated using the *Matplotlib* library.

3. RESULTS AND DISCUSSION

3.1 Descriptive statistics and variability of substrate parameters

The biodegradation of LCB substrates by white-rot fungi during SSF is a fundamentally dynamic process, characterized by the simultaneous and interdependent modification of numerous physicochemical and biochemical properties. To establish biological ground truth for subsequent benchmarking of NIR prediction accuracy, all substrate parameters were chemically quantified at consistent intervals during the 28-day incubation period across five treatments. The descriptive statistics of these measurements, encompassing all samples over the incubation period, are summarized in Table 1.

The moisture content ranged from 53.56% to 65.82% during the 28-day incubation period. In contrast to moisture content, pH showed the highest relative variability among the measured parameters, with a CV of 16.51%, spanning 5.48-9.85. The protein ranged from 6.95% to 9.01%, showing a moderate yet consistent increase during fermentation. The fiber fractions showed distinct yet coordinated ranges. NDF ranged from 46.40% to 54.82%, while ADF fluctuated from 42.82% to 55.98%. ADL, expressed as a concentration (g per 100 g dry matter), ranged from 5.41% to 9.32% with a CV of 12.22%, the second-highest after pH. Because ADL is a concentration relative to the remaining dry matter, changes over time reflect the combined effect of lignin dynamics and

preferential loss of other dry-matter fractions, and should not be read directly as lignin degradation. The cellulose content showed a wider range (35.74%–49.47%), indicating varying accessibility and subsequent degradation following lignin removal. Gross energy was the most consistent measure, ranging from 16.22 to 17.34 MJ/kg DM with a CV of 1.68%.

Table 1. Descriptive statistics of the actual measurement results for all samples over a 28-day incubation period

Parameters	n	Min.	Max.	Mean	SD	CV
Moisture (%)	60	53.56	65.82	58.36	2.48	4.25
pH	60	5.48	9.85	7.23	1.19	16.51
Gross energy (MJ/kg DM)	60	16.22	17.34	16.79	0.28	1.68
Protein (%)	60	6.95	9.01	8.00	0.56	7.02
NDF (%)	60	46.40	54.82	50.99	2.14	4.21
ADF (%)	60	42.82	55.98	50.96	2.94	5.76
ADL (%)	60	5.41	9.32	7.36	0.90	12.22
Cellulose (%)	60	35.74	49.47	43.60	3.13	7.19

Note: n: number of sample datasets; Min: minimum; Max: maximum; SD: standard deviation; CV: coefficient of variation; NDF: neutral detergent fiber; ADF: acid detergent fiber; ADL: acid detergent lignin.

3.2 Dynamic interrelationships during solid-state fermentation biodegradation

To further validate the rationality of the biodegradation, Pearson correlation was performed among all substrate parameters (Figure 1). The results showed that incubation time was significantly positively correlated with moisture content, pH, crude protein (CP), and ADL, indicating that time-dependent substrate alteration was a consistent feature of the SSF process [37]. Over the 28-day incubation, changes in moisture content likely reflect the combined effects of fungal metabolism and evaporative losses inherent to the SSF tray system [38]. Moisture regulation represents one of the most critical yet challenging parameters to sustain in SSF, as white-rot fungi require a substrate moisture range of 60–80% [39] to maintain sufficient water activity for enzymatic catalysis while preserving interparticle porosity essential for oxygen diffusion [40]. Substrate pH also exhibited significant associations with fermentation time, moisture content, protein, and ADL. Although enzyme activities were not directly measured in the present study, these associations are broadly consistent with the reported pH optima of major ligninolytic enzymes. LiP and MnP exhibit peak activity at pH 3.0–4.5 [41], while laccases operate optimally within the pH 5.0–6.0 range [42]. Consequently, the extent of ADL removal was, in part, contingent upon whether ambient pH conditions remained within the functional window of the predominant enzymatic machinery of each strain. The significant correlation between pH and protein is similarly interpretable, with accumulating fungal biomass as metabolic by-products that can buffer the substrate, establishing dynamic feedback between biomass accrual and pH trajectory [43].

Protein, especially in its crude form (CP), was significantly correlated with both time and ADL. This relationship may reflect the simultaneous occurrence of several processes during fungal growth, including the progressive accumulation of fungal mycelium, which is inherently protein-rich relative to the lignocellulosic substrate [43], and the enzymatic mobilization of nitrogen compounds from the substrate matrix as structural carbohydrates and lignin are degraded [44]. The association between CP and ADL may reflect a biologically coherent mechanism in which lignin depolymerization simultaneously exposes bound nitrogen fractions, driving the nutritional upgrading of lignocellulosic substrates that underpins their valorisation as ruminant feed. However, neither fungal biomass nor nitrogen mobilization was directly quantified in this study; these processes should be considered plausible explanations for the observed compositional changes rather than confirmed mechanisms.

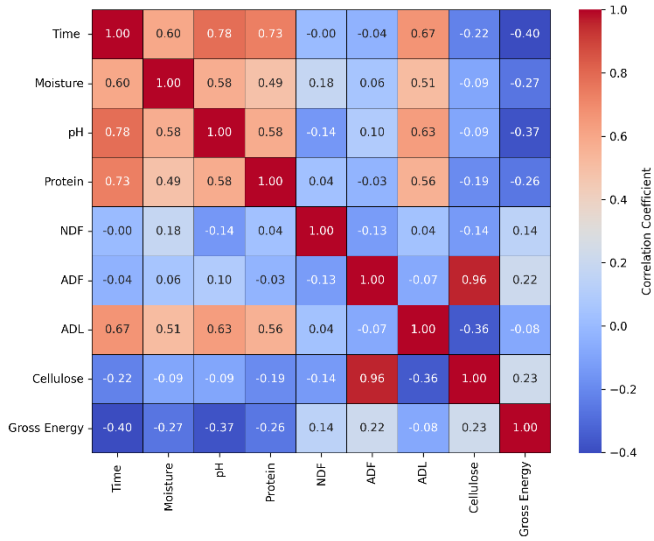

Figure 1. Heatmap of correlation coefficients showing the relationships among incubation time, moisture content, pH, gross energy, crude protein (CP), and fiber fractions (neutral detergent fiber (NDF)), acid detergent fiber (ADF), acid detergent lignin (ADL), and cellulose) during the 28-day solid-state fermentation (SSF) period

Note: Color intensity indicates the strength and direction of the correlation, with red and blue representing positive and negative correlations, respectively.

Among all fiber fractions quantified, ADL exhibited the most extensive correlation network, being significantly associated with fermentation time, moisture content, pH, and CP. As ADL is measured relative to the current dry-matter pool, a positive association with incubation time is consistent with preferential depletion of non-lignin fractions (soluble carbohydrates, hemicellulose, and cellulose), which increases the relative lignin concentration; it does not by itself demonstrate net lignin removal. This pattern is consistent with ADL's recognized role as the primary and most recalcitrant target of fungal enzymatic activity [45], in which lignin depolymerization by LiP and laccases constitutes the rate-limiting step in SSF-mediated substrate valorisation [46]. Because strain-specific effects were not formally tested with a treatment × time analysis, we do not draw conclusions about differences between fungi in ADL dynamics. The sensitivity

of ADL degradation to both moisture-mediated diffusion constraints and pH-induced shifts in enzyme optima is evidenced by its significant correlations with these parameters, collectively illustrating the complexity of the fermentation system. This multidimensional interdependency among substrate parameters provides a compelling rationale for the application of NIRS as an analytical platform, enabling non-destructive, simultaneous, multiparameter characterization from a single spectral acquisition, while concurrently highlighting the analytical challenge that constitutes the focus of this study.

3.3 Analysis of spectral characteristic

Figure 2 presents the mean NIR spectra (1000–2500 nm) for each combination of incubation day (7, 14, 21, and 28 days) and fungal treatment (no strain = control, strain A = *P. chrysosporium*, strain B = *P. ostreatus*, strain C=*T. viride*, and strain D = *L. edodes*), yielding 20 spectral curves of the substrate across the biodegradation process. All spectra shared a common underlying profile shaped by the chemical composition of the LCB substrate matrix. The spectrum was dominated by three main absorption regions that were evident across all treatment groups.

The first is located near 1450 nm, where the first overtone of the O–H stretching vibration is present [47]. The second major feature is a dominant absorption around 1940 nm, corresponding to the combination band of O–H stretching and H–O–H deformation [48], which forms the steepest and strongest absorption peak in the spectrum. The third analytically interesting region covers the range approximately 2100–2350 nm, where overlapping combination bands from C–H deformation, N–H bending, C–O–C ring vibration, and C–H stretching create a broad and complex absorption field [1, 49]. In addition, a minor shoulder near 1200 nm is reliably observed, which can be attributed to the second harmonic of O–H stretching, and a secondary absorption feature near 1750 nm reflects the first harmonic of C–H stretching [50].

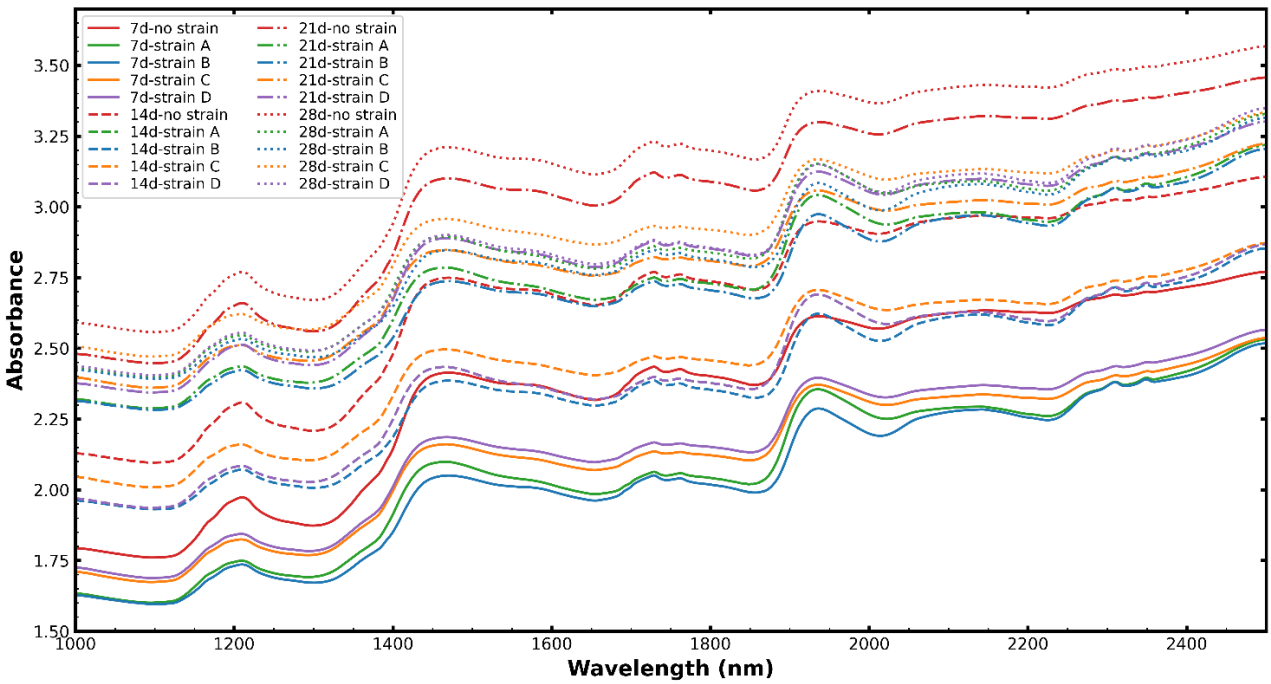

Figure 2. Spectral curves of substrates, averaged by incubation time and across different strains
Note: The colors and shapes of the lines indicate the strains (no strain = control, strain A = *P. chrysosporium*, strain B = *P. ostreatus*, strain C = *T. viride*, and strain D = *L. edodes*) and incubation time (7, 14, 21, and 28 days).

The most distinctive feature of the average spectra in Figure 2 is the systematic increase in relative intensity across the full spectral baseline as incubation time increases, which is consistently observed at all wavelengths from 1000 to 2500 nm. The increase in the spectral baseline is characteristic of increased light scattering within the substrate matrix and can be mechanistically attributed to physical changes occurring during SSF, including changes in surface density and substrate particle geometry resulting from progressive fungal mycelium colonization. These changes increase diffuse reflection scattering, thereby raising the apparent absorbance baseline. This baseline shift is a classic manifestation of the multiplicative scattering component, which needs Multiplicative Scattering Correction (MSC) or Standard Normal Variation (SNV) pre-processing prior to calibration model development [51, 52].

3.4 Important wavelengths selected by Successive Projections Algorithm

Using the SPA, the important wavelengths were selected from the original 1000-2500 nm range, and the results are shown in Table 2. The SPA achieves substantial dimensionality reduction by condensing the 1,557 native spectral channels (equally spaced in wavenumber; non-uniformly spaced in wavelength, see Section 2.4 into 12–21 wavelengths depending on the parameter, a reduction of 98.65–99.23%. This reduction is consistent with previous studies reporting the SPA in NIRS analyses of enzymatic activity and carbohydrate during the saccharification of agricultural materials [27].

Moisture content requires the fewest number of wavelengths (n = 12), consistent with the simple absorption characteristics of water in the NIR region, which are

dominated by the O–H band. Conversely, gross energy requires the most wavelengths (n = 21) because of its narrow range of values (CV = 1.68%), thus requiring more variables to capture subtle spectral differences. Other parameters, such as pH, protein, NDF, ADF, ADL, and cellulose, require 13–16 wavelengths, reflecting moderate spectral complexity. The analysis of wavelength overlap in the eight SPA models indicates a hierarchical structure in the distribution of spectral information (Figure 3).

Table 2. The most important wavelengths (nm) were selected by the Successive Projections Algorithm (SPA) for model development

Parameters	n	Selected Wavelengths	Wavelength Reduction (%)
Moisture	12	1183, 1186, 1187, 1195, 1220, 1238, 1373, 1488, 2038, 2046, 2201, 2245	99.23
pH	13	1000, 1184, 1187, 1191, 1220, 1269, 1361, 1492, 1994, 2045, 2048, 2152, 2247	99.17
Gross energy	21	1012, 1103, 1184, 1185, 1186, 1187, 1189, 1190, 1191, 1192, 1193, 1218, 1229, 1234, 1334, 1469, 1737, 1994, 2043, 2096, 2243	98.65
Protein	14	1000, 1103, 1183, 1187, 1189, 1225, 1228, 1232, 1315, 1887, 1984, 2045, 2048, 2241	99.10
NDF	16	1007, 1180, 1186,1187, 1189, 1191, 1192, 1218, 1234, 1365, 1486, 1832, 1994, 2042, 2179, 2247	98.97
ADF	16	1000, 1184, 1187,1188, 1189, 1190, 1193, 1228, 1238, 1330, 1486, 1984, 2045, 2048, 2241, 2280	98.97
ADL	14	1046, 1183, 1184, 1187, 1191, 1199, 1232, 1238, 1387, 1486, 1994, 2045, 2212, 2245	99.10
Cellulose	16	1005, 1185, 1187, 1189, 1190, 1191, 1218, 1238, 1240, 1363, 1469, 1619, 1994, 2042, 2181, 2241	98.97

Note: n: number of selected wavelengths; NDF: neutral detergent fiber; ADF: acid detergent fiber; ADL: acid detergent lignin.

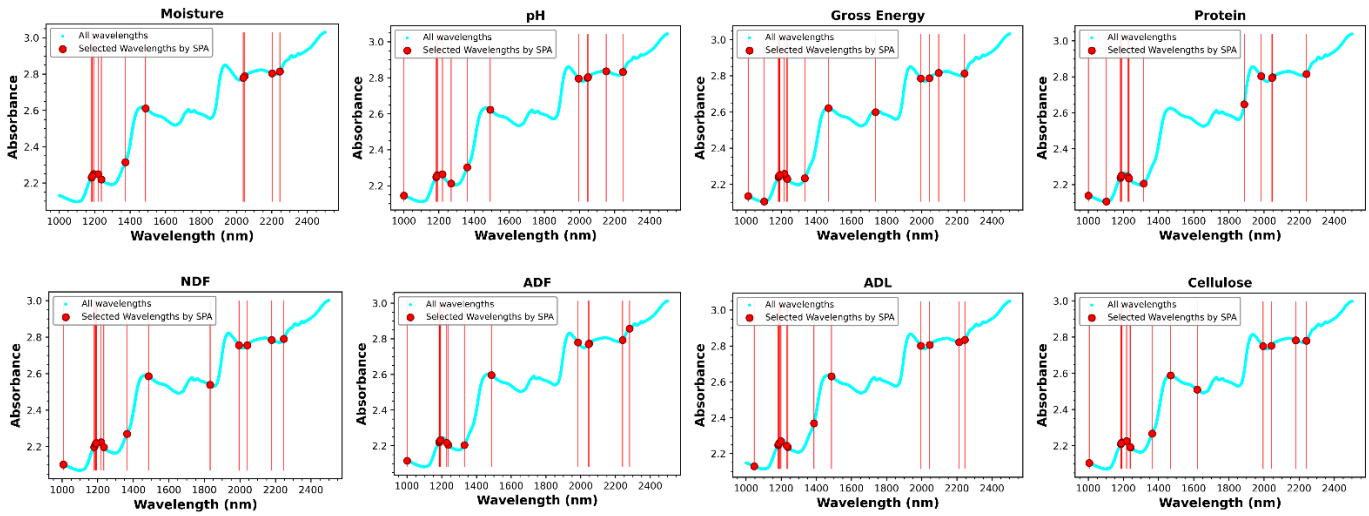

Figure 3. Distribution of Successive Projections Algorithm (SPA)-selected wavelengths used to predict moisture content, pH, gross energy, protein, and fiber fractions (neutral detergent fiber (NDF), acid detergent fiber (ADF), acid detergent lignin (ADL), and cellulose)

Note: The blue line shows the average near-infrared (NIR) absorbance spectrum of the samples across wavelengths from 1000 to 2500 nm. The red markers indicate the wavelengths identified by the SPA algorithm as the most important wavelengths. Selected wavelengths are reported in nm (rounded to the nearest nm); the underlying data are equally spaced in wavenumber (≈3.86 cm^{-1}), so nm intervals between adjacent channels vary across the range.

At the most general level, wavelength clusters around 1186–1193 nm and 1994–2048 nm were consistently selected across nearly all parameters, indicating that these regions captured the most informative variation in the dataset. This probably reflects the dominant influence of moisture dynamics and structural polysaccharide transformations during SSF. These spectral regions encode composite structural information from the lignocellulosic matrix, which correlates with multiple substrate properties, as all measured parameters (moisture, pH, protein, NDF, ADF, ADL, cellulose, and gross energy) are interrelated through fungal biodegradation processes (Figure 1). Spectrally, absorption near 1200 nm is associated with overlapping O–H vibrations and second-order C–H overtones [53], whereas the ~2030 nm region corresponds to N–H combination bands [50]. At an intermediate level, the clusters at 1469–1492 nm (first O–H overtone) and 2241–2247 nm (C–H/N–H/C–O combinations) were selected by most models (six to seven parameters), demonstrating that these locations capture wide spectral properties connected to substrate alterations during SSF. Conversely, at the most detailed level, specific wavelengths were distinctly linked to particular parameters, such as 1737 nm with gross energy, 1887 nm with protein, 1046 and 2212 nm with ADL, and 1619 nm with cellulose. These parameter-specific wavelengths reflect distinct analyte absorptions that are not masked by the general spectral variability of the SSF. Despite the correlation among the parameters, these NIR spectra preserve analyte-specific chemical information, enabling simultaneous and independent prediction of several parameters.

3.5 Machine learning model performance on Successive Projections Algorithm-selected wavelengths

Five ML algorithms, namely PLSR, RF, SVR, AdaBoost, and GBR, were developed and evaluated using a subset of wavelengths selected by SPA to predict eight substrate parameters. Model performance was assessed using R^2_C and R^2_V and the corresponding RMSE values ($RMSE_C$ and $RMSE_V$). The distinctions between R^2_C and R^2_V, as well as the $RMSE_V/RMSE_C$ ratio, were used as indicators to detect overfitting. The complete comparison results are listed in Table 3.

Table 3. Comparative performance of machine learning (ML) models in predicting moisture content, pH, gross energy, protein, and fiber fractions (NDF, ADF, ADL, and cellulose)

Parameters	Method	Hyper-Parameter	Calibration		Validation	
			R^2_C	$RMSE_C$	R^2_V	$RMSE_V$
Moisture (%)	PLSR	LVs = 7	0.51	1.75	0.52	1.58
	RF*	**n_{trees} = 11**	**0.74**	**1.28**	**0.80**	**1.02**
	SVR	C = 3.30, γ = 0.55, ε = 0.30	0.31	2.08	0.33	1.87
	AdaBoost	η = 1, n_{est} = 5	0.95	0.53	0.31	1.90
	GBR	n_{est} = 50, η = 0.09, d_{max} = 5, λ = 5	0.86	0.94	0.32	1.88
pH	PLSR	LVs = 10	0.61	0.73	0.37	0.98
	RF*	**n_{trees} = 7**	**0.74**	**0.60**	**0.73**	**0.63**
	SVR	C = 3.30, γ = 0.55, ε = 0.30	0.64	0.70	0.46	0.91
	AdaBoost	η = 1, n_{est} = 5	0.86	0.44	0.48	0.88
	GBR	n_{est} = 50, η = 0.09, d_{max} = 6, λ = 5	0.84	0.47	0.34	1.00
Gross energy (MJ/kg DM)	PLSR	LVs = 10	0.34	0.23	-0.35	0.31
	RF*	**n_{trees} = 10**	**0.81**	**0.12**	**0.78**	**0.13**
	SVR	C = 3.30, γ = 0.55, ε = 0.30	0.51	0.20	0.31	0.22
	AdaBoost	η = 1, n_{est} = 5	0.91	0.09	0.59	0.17
	GBR	n_{est} = 50, η = 0.09, d_{max} = 6, λ = 5	0.90	0.09	0.68	0.15
Protein (%)	PLSR	LVs = 10	0.56	0.36	0.56	0.41
	RF*	**n_{trees} = 10**	**0.74**	**0.27**	**0.73**	**0.32**
	SVR	C = 3.30, γ = 0.55, ε = 0.30	0.58	0.35	0.67	0.36
	AdaBoost	η = 1, n_{est} = 5	0.86	0.20	0.43	0.47
	GBR	n_{est} = 50, η = 0.09, d_{max} = 6, λ = 5	0.86	0.20	0.68	0.35
NDF (%)	PLSR	LVs = 10	0.21	1.85	-0.15	2.48
	RF	n_{trees} = 5	0.70	1.14	-0.48	2.81
	SVR	C = 3.30, γ = 0.55, ε = 0.30	0.37	1.65	-0.30	2.63
	AdaBoost	η = 1, n_{est} = 5	0.94	0.51	-0.68	3.00
	GBR*	**n_{est} = 50, η = 0.09, d_{max} = 6, λ = 5**	**0.90**	**0.66**	**-0.34**	**2.67**
ADF (%)	PLSR	LVs = 10	0.21	2.66	-0.55	3.17
	RF	n_{trees} = 10	0.61	1.86	-1.89	4.33
	SVR	C = 3.30, γ = 0.55, ε = 0.30	0.33	2.45	-0.40	3.02
	AdaBoost	η = 1, n_{est} = 5	0.95	0.70	-1.86	4.31
	GBR*	**n_{est} = 50, η = 0.09, d_{max} = 6, λ = 5**	**0.88**	**1.02**	**-1.79**	**4.26**
ADL (%)	PLSR	LVs = 9	0.42	0.68	0.54	0.62
	RF	n_{trees} = 9	0.78	0.42	0.72	0.48
	SVR	C = 1.50, γ = 0.20, ε = 0.63	0.57	0.58	0.58	0.59
	AdaBoost	η = 1, n_{est} = 14	0.92	0.25	0.44	0.68
	GBR*	**n_{est} = 6, η = 0.91, d_{max} = 6, λ = 2**	**0.94**	**0.22**	**0.71**	**0.49**
Cellulose (%)	PLSR	LVs = 10	0.20	2.85	-0.24	3.08
	RF	n_{trees} = 10	0.68	1.81	-0.07	2.87
	SVR	C = 3.30, γ = 0.55, ε = 0.30	0.37	2.53	-0.41	3.29
	AdaBoost	η = 1, n_{est} = 5	0.82	1.37	-0.17	2.99
	GBR*	**n_{est} = 50, η = 0.09, d_{max} = 6, λ = 5**	**0.86**	**1.21**	**-0.17**	**3.00**

Note: ***** denotes the best-performing model, NDF: neutral detergent fiber; ADF: acid detergent fiber; ADL: acid detergent lignin; R^2_C: coefficient of determination for calibration; $RMSE_C$: root mean square error of calibration; R^2_V: coefficient of determination for validation; $RMSE_V$: root mean square error of validation; LVs: latent variables; n_{trees}: n_trees; C: regularization parameter, γ: kernel coefficient, ε: epsilon-insensitive loss parameter; η: learning _rate; n_{est}: n_estimators; d_{max}: max_depth; λ: L2 regularization parameter; PLSR: Partial Least Squares Regression; RF: Random Forest; SVR: Support Vector Regression; AdaBoost: Adaptive Boosting; GBR: Gradient Boosting Regression.

The analysis revealed two significant results. First, the moisture content, pH, gross energy, protein, and ADL can be predicted with adequate accuracy using NIRS based on the selected SPA wavelength. Among the five algorithms evaluated, the RF model consistently demonstrated the strongest validation performance for several substrate parameters, yielding the highest R^2_V values for moisture content (R^2_V = 0.80; $RMSE_V$ = 1.02%), pH (R^2_V = 0.73; $RMSE_V$ = 0.63), gross energy (R^2_V = 0.78; $RMSE_V$ = 0.13 MJ/kg DM), and protein (R^2_V = 0.73; $RMSE_V$ = 0.32%), while, co-achieving the best performance for ADL (R^2_V = 0.72; $RMSE_V$ = 0.48%). The differences between the R^2_C and R^2_V values for the moisture content, pH, gross energy, CP, and ADL were minimal, at −0.06, 0.01, 0.03, 0.01, and 0.06, respectively. These values were close to zero, indicating that the RF model provided comparatively stable predictions for the parameters with satisfactory validation performance. Figure 4 depicts scatter plots of measured versus predicted values, illustrating the calibration and validation of the best-performing models for the eight substrate parameters.

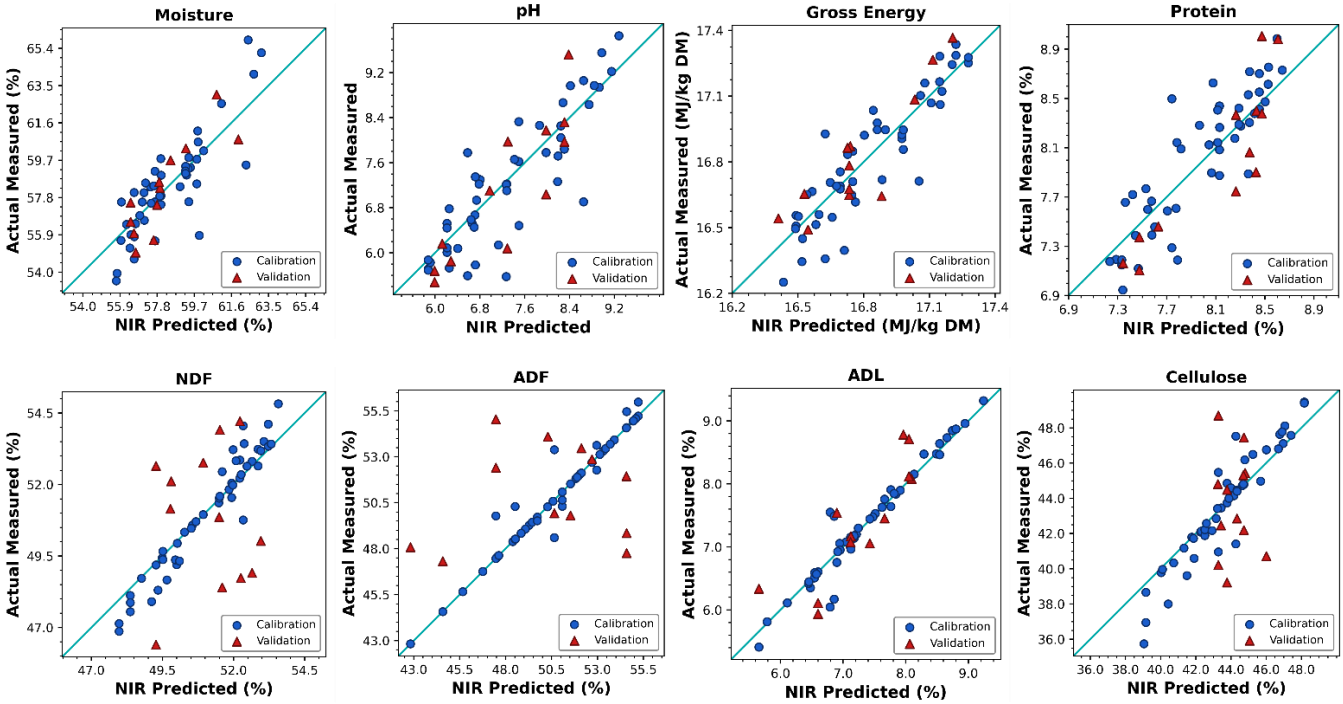

Figure 4. Scatter plots comparing measured and predicted values for moisture content, pH, gross energy, protein, and fiber fractions (neutral detergent fiber (NDF), acid detergent fiber (ADF), acid detergent lignin (ADL), and cellulose) using the best-performing machine learning (ML) models. RF provided optimal validation performance for moisture content, pH, gross energy, and crude protein (CP), whereas GBR showed the highest validation performance among the evaluated models for NDF, ADF, ADL, and cellulose

The RF models for these five parameters fell into the approximate quantification category (R^2_V = 0.70–0.89), which signifies adequate practical value for process monitoring [34]. This is particularly beneficial when a rough estimate of a parameter is sufficient to initiate corrective measures if the actual value deviates significantly from the desired range of values. Moisture content serves as a compelling illustration of this case. In SSF, optimal moisture content typically ranges from 60–70%, which is critical for efficient ADL degradation, as it ensures proper fungal growth, nutrient diffusion, and enzyme activity, while preventing oxygen limitation [5, 54]. Although NIRS estimates are approximate, they can effectively identify deviations from this optimal range, allowing for timely adjustments in moisture or aeration to prevent disruptions in the SSF process [55]. It should be emphasized that ADL (the goals of this study) was successfully predicted with adequate accuracy.

This achievement has scientific significance that extends beyond technical validation, as lignin degradation is a central goal of white-rot fungal fermentation [56]. Furthermore, among the parameters examined in this study, ADL demonstrated a strong association with fermentation time, moisture content, pH, and protein (as illustrated in Figure 1). This relationship highlights ADL's function as a key factor that comprehensively reflects the biochemical processes of the substrate during the fungal biodegradation process. Thus, the ability of NIRS to simultaneously monitor ADL, moisture content, pH, and protein underscores its potential as a comprehensive and reliable diagnostic tool for real-time monitoring in the management of lignocellulose fermentation processes. By identifying which parameters are suitable for real-time monitoring and which require further methodological refinement, this study lays a solid foundation for the development of NIRS-based process analytical technology in SSF systems.

Among the five algorithms assessed, the RF algorithm was the sole method that consistently demonstrated the ability to avoid overfitting. In contrast to boosting-based algorithms, such as AdaBoost and GBR, which demonstrated high R^2_C values during the calibration phase, these algorithms encountered difficulties in sustaining their performance during the validation phase across most parameters. This pattern arises from the fundamental differences between boosting and bagging strategies. The generalization advantage of RF stems from its construction strategy. RF creates a large number of uncorrelated decision trees through bootstrap aggregation (bagging) and random feature sampling at each split point [57, 58]. This approach results in a model with lower variance than any individual tree, providing more stable predictions and reducing the risk of overfitting owing to noise [57]. In the SPA-reduced spectral space of 12–21 wavelengths, the risk of modelling spurious collinearity is significantly lower compared to full-spectrum PLSR. Consequently, the RF ensemble structure is well-suited to capturing nonlinear relationships and interaction-based dynamics between NIR reflectance and the chemical attributes of the substrate during biodegradation.

The second finding of this study reveals a complete failure of NIRS predictions for NDF, ADF, and cellulose, all of which fell below the minimum threshold deemed acceptable for practical applications (R^2_V < 0.50) [35]. All five algorithms produced negative R^2_V values for these parameters, with $RMSE_V$ values ranging from 2.48 to 4.33, far exceeding the $RMSE_C$ values, which ranged from 0.51 to 1.86 for the most optimal model. The $RMSE_V/RMSE_C$ ratio of 6.16 for AdaBoost on ADF illustrates significantly poorer validation

performance than calibration (Table 3 and Figure 4). The contributing factors to this failure include a small dataset size (n = 60) and substantial spectral overlap in the NIR range among NDF, ADF, and cellulose [59]. Structurally, ADF is a subset of NDF, and cellulose is the main component of ADF [60], resulting in shared NIR absorption characteristics primarily influenced by O–H, C–H, and C–O–C vibrations from the polysaccharide backbone [61]. As shown in Table 2, the wavelengths selected by SPA for NDF and ADF overlap significantly, with 14 of the 16 variables being identical. This suggests that the models were trained on nearly identical spectral inputs for predicting distinct reference values. Therefore, a larger calibration dataset is essential to effectively differentiate parameters with high intercorrelation, including NDF, ADF, and cellulose.

Several limitations of this study should be considered when interpreting the results. First, baseline (day-0) measurements were not collected, and neither dry-matter loss nor the optimal treatment duration was determined during fermentation. Consequently, absolute lignin loss and protein enrichment could not be quantified, and the optimal treatment duration remains unidentified from the available data. All compositional results are therefore presented as concentrations along with relative changes over the sampled interval (days 7–28). Addressing this limitation would necessitate a treatment × time mixed-model analysis that includes baseline (day-0) sampling and dry-matter accounting, which is identified as a priority for future research. Second, the dataset is limited in size (n = 60, of which only 45 samples were available for calibration) and was derived from a single fermentation run conducted under uncontrolled ambient temperature (37 ± 3 °C) without continuous environmental monitoring. A calibration set of this size limits the robustness of the ML models evaluated, particularly the ensemble-based AdaBoost and GBR algorithms, which are more prone to overfitting under limited sample sizes. Thus, the resulting models should be considered as a preliminary internal validation rather than as widely applicable predictive tools. Third, each sample was scanned as a single, intact packed cup, meaning that within-sample heterogeneity was not assessed.

4. CONCLUSIONS

This study demonstrates that NIRS combined with SPA and ML enables reliable prediction of selected compositional parameters during LCB biodegradation, particularly moisture, pH, protein, and ADL. However, consistent prediction failure for NDF, ADF, and cellulose highlights intrinsic limitations due to spectral overlap and structural interdependence. Among the ML models evaluated, the RF model showed the best predictive performance, with R^2_V values ranging from 0.71 to 0.80 for moisture (R^2_V = 0.80; $RMSE_V$ = 1.02%), pH (R^2_V = 0.73; $RMSE_V$ = 0.63), gross energy (R^2_V = 0.78; $RMSE_V$ = 0.13 MJ/kg DM), protein (R^2_V = 0.73; $RMSE_V$ = 0.32%), and ADL (R^2_V = 0.72; $RMSE_V$ = 0.48%). Furthermore, the RF model exhibited minimal calibration-validation deviation, indicating satisfactory robustness and a low tendency for overfitting. Nevertheless, none of the evaluated ML algorithms achieved satisfactory generalization performance for NDF, ADF, or cellulose, as reflected by negative R^2_V values. The consistently poor predictive performance observed across all ML models indicates that structurally complex carbohydrate fractions remain difficult to predict using the current NIRS dataset. This limitation is likely attributable to multiple interacting factors, including the limited sample size, spectral overlap among lignocellulosic constituents, moisture-induced spectral interference, uncertainty in the reference chemical analyses, and the current feature-selection and validation strategies, rather than to an inherent limitation of NIRS itself. However, the integration of NIRS with ML provides an efficient approach for monitoring key biodegradation indicators, particularly moisture, pH, protein, and ADL during SSF. Future research should therefore focus on expanding datasets, improving spectral preprocessing, and integrating complementary analytical techniques to enhance predictive robustness and generalization.

ACKNOWLEDGMENT

This study was funded by the Doctoral Dissertation Research Grant 2025 from the Ministry of Education, Culture, Research, and Technology – Indonesia (KEMDIKTISAINTEK), as specified in assignment agreement No. 113/C3/DT.05.00/PL/2025.

REFERENCES

[1] Wahyudi, I., Wajizah, S., Munawar, A.A., Kaloudis, E., Samadi, S. (2025). Chemometric approach based on feed NIR spectra for rapid assessment of digestibility profiles in unconventional feedstuffs. BIO Web of Conferences, 206: 02001. https://doi.org/10.1051/bioconf/202520602001

[2] Priyadarshani, S., Nandal, P., Arora, A., et al. (2025). A green process for lignin extraction and lignocellulose degrading enzyme production from rice straw by solid state fermentation with Streptomyces thermoviolaceous strains. Waste Management Bulletin, 3(3): 100207. https://doi.org/10.1016/j.wmb.2025.100207

[3] Fu, R., Han, L., Li, Q., Li, Z., Dai, Y., Leng, J. (2025). Studies on the concerted interaction of microbes in the gastrointestinal tract of ruminants on lignocellulose and its degradation mechanism. Frontiers in Microbiology, 16: 1554271. https://doi.org/10.3389/fmicb.2025.1554271

[4] Wang, Y.L., Wang, W.K., Wu, Q.C., Yang, H.J. (2022). The release and catabolism of ferulic acid in plant cell wall by rumen microbes: A review. Animal Nutrition, 9: 335-344. https://doi.org/10.1016/j.aninu.2022.02.003

[5] Khan, N.A., Sufyan, A., Ullah, M.W., et al. (2025). Fungal pretreatment of lignocellulosic feedstocks: Challenges and opportunities in lignin degradation, structural polysaccharide preservation, and conversion into ruminant feed. Journal of Agricultural and Food Chemistry, 73(44): 27909-27932. https://doi.org/10.1021/acs.jafc.5c09647

[6] Kumar, A., Chandra, R. (2020). Ligninolytic enzymes and its mechanisms for degradation of lignocellulosic waste in environment. Heliyon, 6(2): e03170. https://doi.org/10.1016/j.heliyon.2020.e03170

[7] Benavides, V., Ciudad, G., Pinto-Ibieta, F., Robledo, T., Rubilar, O., Serrano, A. (2024). Enhancing laccase and manganese peroxidase activity in white-rot fungi: The role of copper, manganese, and lignocellulosic

substrates. Agronomy, 14(11): 2562. https://doi.org/10.3390/agronomy14112562
[8] Mohd Din, A.R.J., Chua, T.L., Mohamad Azam, Z., Othman, N.Z., El-Enshasy, H.A., Yong, J.W.H. (2025). Solid-state fermentation of spent mushroom substrate through a synergistic fungal consortium for enzymatic cocktail production. Bioresource Technology Reports, 32: 102417. https://doi.org/10.1016/j.biteb.2025.102417
[9] Mulatu, A., Alemu, T., Megersa, N., Vetukuri, R.R. (2021). Optimization of culture conditions and production of bio-fungicides from Trichoderma species under solid-state fermentation using mathematical modeling. Microorganisms, 9(8): 1675. https://doi.org/10.3390/microorganisms9081675
[10] Nayan, N., Sonnenberg, A.S.M., Hendriks, W.H., Cone, J.W. (2017). Differences between two strains of Ceriporiopsis subvermispora on improving the nutritive value of wheat straw for ruminants. Journal of Applied Microbiology, 123(2): 352-361. https://doi.org/10.1111/jam.13494
[11] Van Kuijk, S.J.A., Sonnenberg, A.S.M., Baars, J.J.P., Hendriks, W.H., Cone, J.W. (2015). Fungal treated lignocellulosic biomass as ruminant feed ingredient: A review. Biotechnology Advances, 33(1): 191-202. https://doi.org/10.1016/j.biotechadv.2014.10.014
[12] Abbasi, V., Sar, T., Agnihotri, S., Mahboubi, A. (2026). Valorization of stale bread through solid state fermentation using Neurospora intermedia. Waste and Biomass Valorization, pp. 1-13. https://doi.org/10.1007/s12649-026-03552-9
[13] Mao, L., Sonnenberg, A.S.M., Hendriks, W.H., Cone, J.W. (2017). Preservation of Ceriporiopsis subvermispora and Lentinula edodes treated wheat straw under anaerobic conditions. Journal of the Science of Food and Agriculture, 98(3): 1232-1239. https://doi.org/10.1002/jsfa.8745
[14] Van Soest, P.J., Robertson, J.B., Lewis, B.A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. Journal of Dairy Science, 74(10): 3583-3597. https://doi.org/10.3168/jds.s0022-0302(91)78551-2
[15] Wahyudi, I., Wajizah, S., Munawar, A.A., Samadi. (2025). NIRS-based rapid analysis of crude fiber and crude protein in fermented citronella residues as non-conventional feedstuffs. IOP Conference Series: Earth and Environmental Science, 1476(1): 012020. https://doi.org/10.1088/1755-1315/1476/1/012020
[16] Olagunju, O., Stump, M., Li, Y. (2025). Machine learning-enabled nondestructive quality analysis of animal protein-based foods: A comprehensive review. Agricultural Products Processing and Storage, 1(1): 7. https://doi.org/10.1007/s44462-025-00013-9
[17] 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 & Nature and Ecodynamics, 18(4): 951-956. https://doi.org/10.18280/ijdne.180422
[18] 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
[19] Jiang, H., Zhang, H., Chen, Q., Mei, C., Liu, G. (2015). 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
[20] Bearzi, G., Grassi, S., Cappa, C., Rollini, M., Alamprese, C. (2026). Innovative monitoring of medicinal mushroom growth on food by-product substrates in solid-state fermentation. Food Bioscience, 79: 108849. https://doi.org/10.1016/j.fbio.2026.108849
[21] Akbar, F.N., Mahmood, S., Mueen-ud-din, G., Yamin, M., Murtaza, M.A. (2024). Exploring the effects of drying method and temperature on the quality of dried basil (Ocimum basilicum L.) leaves: A sustainable and eco-friendly drying solution. Resources, 13(9): 121. https://doi.org/10.3390/resources13090121
[22] 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
[23] AOAC. (2016). Official Methods of Analysis of the Association of Official Analytical Chemists, 20th ed. Washington, DC: AOAC Inc.
[24] 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: 20-28. https://doi.org/10.1016/j.aca.2019.08.067
[25] Ye, S., Wang, D., Min, S. (2008). Successive projections algorithm combined with uninformative variable elimination for spectral variable selection. Chemometrics and Intelligent Laboratory Systems, 91(2): 194-199. https://doi.org/10.1016/j.chemolab.2007.11.005
[26] Li, B., Martin, E., Morris, J. (2001). Latent variable selection in partial least squares modelling. IFAC Proceedings Volumes, 34(25): 463-468. https://doi.org/10.1016/s1474-6670(17)33867-3
[27] He, H.J., Xing, L., Lü, M., et al. (2025). Machine learning-assisted micro NIR for rapidly monitoring the variations of enzymatic activity and carbohydrates during saccharification of stored sweet potatoes. LWT, 232: 118409. https://doi.org/10.1016/j.lwt.2025.118409
[28] 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 & Nature and Ecodynamics, 20(11): 2675-2686. https://doi.org/10.18280/ijdne.201119
[29] Peng, H., Yi, L., Fan, X., Zhang, J., Gu, Y., Wang, S. (2025). Near-infrared spectroscopy assisted by random forest for predicting the physicochemical indicators of yak milk powder. Food Chemistry, 478: 143555. https://doi.org/10.1016/j.foodchem.2025.143555
[30] Zhu, R., Wu, X., Wu, B., Gao, J. (2024). High-accuracy classification and origin traceability of peanut kernels based on near-infrared (NIR) spectroscopy using Adaboost - maximum uncertainty linear discriminant analysis. Current Research in Food Science, 8: 100766. https://doi.org/10.1016/j.crfs.2024.100766
[31] Zheng, R., Jia, Y., Ullagaddi, C., et al. (2024). Optimizing feature selection with gradient boosting

machines in PLS regression for predicting moisture and protein in multi-country corn kernels via NIR spectroscopy. Food Chemistry, 456: 140062. https://doi.org/10.1016/j.foodchem.2024.140062
[32] 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 & Nature and Ecodynamics, 19(2): 379-386. https://doi.org/10.18280/ijdne.190203
[33] Williams, P., Norris, K.H. (1987). Near-infrared technology in the agricultural and food industries. American Association of Cereal Chemists.
[34] 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
[35] 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
[36] 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
[37] Zhao, X., Wang, F., Fang, Y., et al. (2020). High-potency white-rot fungal strains and duration of fermentation to optimize corn straw as ruminant feed. Bioresource Technology, 312: 123512. https://doi.org/10.1016/j.biortech.2020.123512
[38] Rodríguez, L.A., Groff, M.C., Garay, S.A., Díaz, M.E., Sardella, M.F., Scaglia, G. (2025). Bioconversion of olive pomace: A solid-state fermentation strategy with Aspergillus sp. for detoxification and enzyme production. Fermentation, 11(8): 456. https://doi.org/10.3390/fermentation11080456
[39] Onu Olughu, O., Tabil, L.G., Dumonceaux, T., Mupondwa, E., Cree, D. (2022). Optimization of solid-state fermentation of switchgrass using white-rot fungi for biofuel production. Fuels, 3(4): 730-752. https://doi.org/10.3390/fuels3040043
[40] Perwez, M., Al Asheh, S. (2025). Valorization of agro-industrial waste through solid-state fermentation: Mini review. Biotechnology Reports, 45: e00873. https://doi.org/10.1016/j.btre.2024.e00873
[41] Susanti, E., Ardyati, T., Suharjono, S., Aulanni'am, A. (2022). Characterization of lignin peroxidase from the suspected novel strain Phanerochaete chrysosporium ITB isolate. Indonesian Journal of Chemistry, 22(5): 1386. https://doi.org/10.22146/ijc.74637
[42] Viswanath, B., Rajesh, B., Janardhan, A., Kumar, A.P., Narasimha, G. (2014). Fungal laccases and their applications in bioremediation. Enzyme Research, 2014: 163242. https://doi.org/10.1155/2014/163242
[43] Sar, T., Larsson, K., Fristedt, R., Undeland, I., Taherzadeh, M.J. (2022). Demo-scale production of protein-rich fungal biomass from potato protein liquor for use as innovative food and feed products. Food Bioscience, 47: 101637. https://doi.org/10.1016/j.fbio.2022.101637
[44] Kohl, L., Wanek, W., Keiblinger, K., et al. (2023). Nutrient controls on carbohydrate and lignin decomposition in beech litter. Geoderma, 429: 116276. https://doi.org/10.1016/j.geoderma.2022.116276
[45] Ratni, E., Lendrawati, L., Hefzi, F., Muqarramah, M. (2025). Isolation and identification of local white-rot fungi from West Sumatra and their potential for biodelignification of ruminant feed. Journal of Advanced Veterinary and Animal Research, 12(2): 558. https://doi.org/10.5455/javar.2025.l920
[46] Gokul, T.A., Ramesh Kumar, K., Dang, K.D., et al. (2026). Unlocking lignocellulosic biomass: Biocatalyst-driven strategies for sustainable biofuel generation. Carbon Resources Conversion, p. 100417. https://doi.org/10.1016/j.crcon.2026.100417
[47] Gowen, A.A., Amigo, J.M., Tsenkova, R. (2013). Characterisation of hydrogen bond perturbations in aqueous systems using aquaphotomics and multivariate curve resolution-alternating least squares. Analytica Chimica Acta, 759: 8-20. https://doi.org/10.1016/j.aca.2012.10.007
[48] Ishigaki, M., Koizumi, K., Asano, K., et al. (2026). Gravitational effects on the hydrogen bond network of water and ionic solutions revealed by near infrared spectroscopy under simulated microgravity. Scientific Reports, 16(1): 13497. https://doi.org/10.1038/s41598-026-44169-1
[49] Dai, C., Xu, X., Huang, W., et al. (2022). Monitoring of critical parameters in thermophilic solid-state fermentation process of soybean meal using NIR spectroscopy and chemometrics. Food Measure, 17(1): 576-585. https://doi.org/10.1007/s11694-022-01628-3
[50] Weyer, L.G., Lo, S.C. (2006). Spectra-structure correlations in the near-infrared. In Handbook of Vibrational Spectroscopy, 3: 1817-1837. https://doi.org/10.1002/0470027320.s4102
[51] Dhanoa, M.S., Lister, S.J., Sanderson, R., Barnes, R.J. (1994). The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. Journal of Near Infrared Spectroscopy, 2(1): 43-47. https://doi.org/10.1255/jnirs.30
[52] Wu, Y., Peng, S., Xie, Q., Han, Q., Zhang, G., Sun, H. (2019). An improved weighted multiplicative scatter correction algorithm with the use of variable selection: Application to near-infrared spectra. Chemometrics and Intelligent Laboratory Systems, 185: 114-121. https://doi.org/10.1016/j.chemolab.2019.01.005
[53] He, H.J., Zhang, C., Xing, L., et al. (2025). 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
[54] Yao, Y., Luo, Y., Wu, M., et al. (2025). Contribution of moisture to functional microbial succession and functional expression in medium-high temperature Daqu. Food Bioscience, 69: 106831. https://doi.org/10.1016/j.fbio.2025.106831
[55] Dashtban, M., Schraft, H., Syed, T.A., Qin, W. (2010). Fungal biodegradation and enzymatic modification of lignin. International Journal of Biochemistry and Molecular Biology, 1(1): 36-50.
[56] Bamidele, M.O., Bamikale, M.B., Cárdenas-Hernández, E., et al. (2025). Bioengineering in solid-state fermentation for next sustainable food bioprocessing. Next Sustainability, 6: 100105. https://doi.org/10.1016/j.nxsust.2025.100105

[57] Karabadji, N.E.I., Amara Korba, A., Assi, A., Seridi, H., Aridhi, S., Dhifli, W. (2023). Accuracy and diversity-aware multi-objective approach for random forest construction. Expert Systems with Applications, 225: 120138. https://doi.org/10.1016/j.eswa.2023.120138
[58] Yang, B., Jahed Armaghani, D., Fattahi, H., et al. (2025). Optimized random forest models for rock mass classification in tunnel construction. Geosciences, 15(2): 47. https://doi.org/10.3390/geosciences15020047
[59] Tuite, M., Liu, S., Phippen, W., Starke, M., Chopra, R. (2025). Rapid and non-destructive screening of seed components in domesticated pennycress using near-infrared spectroscopy. Industrial Crops and Products, 235: 121752. https://doi.org/10.1016/j.indcrop.2025.121752
[60] Łopatyńska, A., Wolko, J., Wolko, Ł., Bocianowski, J., Spychała, J., Noweiska, A. (2025). Genome-wide association study for markers related to protein, fiber (ADF and NDF) and oil content in winter oilseed rape seeds (Brassica napus L.). International Journal of Molecular Sciences, 26(24): 11931. https://doi.org/10.3390/ijms262411931
[61] Gieroba, B., Kalisz, G., Krysa, M., Khalavka, M., Przekora, A. (2023). Application of vibrational spectroscopic techniques in the study of the natural polysaccharides and their cross-linking process. International Journal of Molecular Sciences, 24(3): 2630. https://doi.org/10.3390/ijms24032630