Received: 21 November 2025 | Accepted: 20 March 2026

DOI: 10.1002/agj2.70386

Agronomy Journal

ORIGINAL ARTICLE

Agronomy, Soils, and Environmental Quality

Hyperspectral reflectance and machine learning to monitor legume feed quality

Chinmayi H. K.[1] | **K. Colton Flynn**[1] | **Gurjinder S. Baath**[2] | **Trey O. Lee**[1] | **Prasanna Gowda**[3] | **Brian Northup**[4] | **Amanda Ashworth**[5]

[1]USDA-ARS Grassland Soil and Water Research Laboratory, Temple, Texas, USA

[2]Blackland Research & Extension Center, Texas A&M AgriLife Research, Temple, Texas, USA

[3]USDA-ARS Southeast Area, Stoneville, Mississippi, USA

[4]USDA-ARS Grazinglands Research Laboratory, El Reno, Oklahoma, USA

[5]USDA-ARS Poultry Production and Product Safety Research, O-303 Poultry Science Center, Fayetteville, Arkansas, USA

Correspondence
K. Colton Flynn, USDA-ARS Grassland Soil and Water Research Laboratory, 808 E. Blackland Road, Temple, TX 76502, USA.
Email: colton.flynn@usda.gov

Assigned to Associate Editor Yuanshuo Qu.

Funding information
Agricultural Research Service

Abstract

Hyperspectral remote sensing is transforming precision agriculture by capturing detailed spectral information that supports improved crop management. This study evaluated forage quality in soybean (*Glycine max*), tepary bean (*Phaseolus acutifolius*), and moth bean (*Vigna aconitifolia*) using in situ hyperspectral measurements and convolved wavebands to predict neutral detergent fiber, acid detergent fiber, in vitro true digestibility, and crude protein with machine learning models including k-nearest neighbors (k-NN), partial least squares regression (PLSR), support vector networks (SVN), and random decision forests (RDFs). To assess the potential of future hyperspectral satellites, hyperspectral wavelengths were convolved to represent the proposed spectral bands of CHIME. Additional analyses compared models incorporating legume type as a variable with those developed separately for each species. Results indicate that both in situ hyperspectral data and CHIME-simulated bands effectively predicted key forage traits, with RDF and SVN consistently outperforming k-NN and PLSR. SVN achieved the highest predictive accuracy while requiring fewer computational resources, whereas RDF predictions were slightly lower by 1%–6%. In contrast, k-NN performance declined by 18%–25%, and PLSR showed the weakest predictive ability. Among additional model comparisons, it was found that model building of individual legume types generally outperformed models incorporating legume type as a variable. The results also provide insights into the nutritional composition of legume forages that can aid in their selection and utilization in various agronomic applications. Furthermore, these findings provide a foundation for advancements in hyperspectral remote sensing, providing crucial insights for improving precision agriculture practices.

ABBREVIATIONS: ADF, acid detergent fiber; CP, crude protein; DAP, days after planting; FD, first derivative; IVTD, in vitro true digestibility; k-NN, k-nearest neighbors; ML, machine learning; NDF, neutral detergent fiber; PLSR, partial least squares regression; RDF, random decision forest; RMSE, root mean square error; SG, Savitsky–Golay; SVN, support vector networks.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2026 The Author(s). Agronomy Journal published by Wiley Periodicals LLC on behalf of American Society of Agronomy.

Plain Language Summary

Hyperspectral remote sensing is a new technology that helps farmers improve crop production by giving detailed information about plants. In this study, we looked at the forage quality of three legumes: soybean, tepary bean, and moth bean. Using field data, we predicted fiber content, digestibility, and protein levels. To do this, we tested several machine learning methods, including k-nearest neighbors, partial least squares regression, support vector machine, and random forest. We also explored how this technology could work with a future satellite by simulating its spectral bands. Models that treated each legume species separately gave better results than those combining all species. Overall, random forest and support vector machine models performed best, with support vector machine being especially efficient. These findings show that hyperspectral sensing can help farmers measure forage quality and support precision agriculture for producing high-quality crops.

1 | INTRODUCTION

Hyperspectral remote sensing is a technology revolutionizing the field of precision agriculture (Ram et al., 2024). It is an affordable and efficient alternative to traditional biophysical and biochemical analysis methods offering a level of accuracy and spatial detail that was previously impossible to achieve due to intensive labor and expenses (P. Singh et al., 2020). Information from remote sensing measurements can help improve management decisions leading to the optimization of crop yield, enhanced soil quality, and ultimately increased profits (Abuova et al., 2020; Chen et al., 2022; Gopathoti, 2023; Hedley, 2014; McConnell, 2019; Schimmelpfennig, 2017; Shannon et al., 2018; Vatin et al., 2024; Wang et al., 2015).

Space-borne hyperspectral platforms and sensors such as CHIME (Buschkamp et al., 2023), EnMap (Schwieder et al., 2014), and PRISMA (PRecursore IperSpettrale della Missione Applicativa; Cogliati et al., 2021) have potential to provide farm and ranch managers unparalleled access to highly accurate spatial data for forage quality, offering an affordable and efficient alternative to traditional methods for optimizing management decisions. Hyperspectral reflectance data can provide valuable information about the nutritional and chemical composition of legumes (Baath, Baath, et al., 2020; Syamsi et al., 2022). Hyperspectral data can also be utilized to accurately predict key forage quality parameters, including crude protein (CP), fiber, and digestibility (Baath, Baath, et al., 2020).

Agriculture is a highly managed and dynamic industry that requires precise data and information to optimize production as well as reduce waste and environmental degradation. Hyperspectral imaging has the potential to provide such information by providing detailed insights into forage health, nutrient levels, and phenology/growth patterns (Lu et al., 2020). Although numerous studies have applied hyperspectral imaging to agricultural production and crop trait estimation, these efforts have largely centered on individual crop species, controlled environments, or simplified management regimes (Guerri et al., 2024; Kumari et al., 2025; Ram et al., 2024). Consequently, there remains a need for research that incorporates diverse crop types and real-world management practices while accounting for advancements in satellite-based hyperspectral sensors. Furthermore, it is crucial for research to consider multiple crop types and management regimes, such as cover cropping, when developing predictive models for the agricultural industry, especially given the growing pressures on food systems and agroecosystem health (Jones et al., 2021).

Perennial warm-season grasses such as old-world bluestems (*Bothriochloa spp.*) and bermudagrass (*Cynodon dactylon*) support many cattle grazing operations in the Southern Great Plains of the United States. These perennials offer abundant fodder for grazing, but their nutritional value frequently diminishes in mid-summer to late summer, leading to ongoing challenges in meeting nutritional requirements for cattle production (Guretzky et al., 2011; Tracy et al., 2010). This decline in nutrients provides an opportunity for legumes to be of greater nutrient value during the summer months in which struggles are experienced (Baath, Baath, et al., 2020).

Legumes are renowned for their nitrogen (N)-fixing ability, whereby they convert atmospheric nitrogen (N) into a form that plants can utilize, improving soil health and fertility (Kebede, 2021). This process also provides a natural and sustainable source of protein for livestock, resulting in legumes serving as a valuable crop in agricultural systems. Additionally, many legumes can produce significant amounts of high nitrogen (N) biomass, which can serve as green manure to supply N to succeeding crops (Talgre et al., 2018). Legumes like soybeans (*Glycine max*) are widely used in US agriculture as rotational or cover crops for maize (*Zea mays*), primarily for their ability to fix N (Ladha et al., 2022). However, additional novel legumes like tepary bean (*Phaseolus acutifolius*)

and moth bean (*Vigna aconitifolia*) offer valuable solutions for meeting cattle grazing nutritional needs and/or cover crop needs while simultaneously promoting agrobiodiversity (Baath et al., 2018; H. Singh et al., 2023).

Increased biodiversity can pose difficulties for hyperspectral remote sensing interventions. Studies have shown that crop texture and growth pattern can notably reduce the precision of models predicting biophysical and biochemical characteristics (Agapiou, 2020). For legumes with varying textures and growth orientations, such as soybean, tepary bean, and moth bean, creating a unified model to accurately predict the biophysical and biochemical characteristics of a diverse range of legume species has the potential to present challenges. Soybean is a commonly cultivated, warm-season legume with an erect growth habit in the Southern Great Plains of the United States (Rao et al., 2005). Tepary bean, which originated in northwestern Mexico and the southwestern United States, has a bush or semi-vine growth habit and is known for its excellent forage and cover capabilities, particularly in heat and drought (Baath, Northup, et al., 2020). Similarly, moth bean, known for its versatile prostrate growth habit, thrives in the dry landscapes of India. It provides nutrient-rich pulses for human diets, serves as forage for livestock, and acts as a cover crop to help retain and protect soil (Baath et al., 2018). These differences in their textures and growth orientations mean that a uniform approach may result in challenges for model building associated with livestock feed quality (Galvão et al., 2018; Flynn et al., 2023). The challenges provide an opportunity to incorporate advanced analytical methodologies such as machine learning (ML).

Precision agriculture management, which involves using data-driven approaches to optimize crop management practices, is one area where hyperspectral remote sensing and ML can make a significant impact on grazing and cover crop systems (Finger et al., 2019; Monteleone et al., 2020). By using remote sensing data to identify variability within a field, farmers can tailor their management practices to the specific needs across spatial scales (Flynn et al., 2020).

ML techniques have already shown great potential in predicting and classifying various aspects of legume crops, such as disease detection (Cho, 2024), yield prediction under climate change (Na & Na, 2024), and understanding the relationships between chemical composition and nutrient fermentability (Ikhwanti et al., 2020). ML algorithms can analyze large datasets encompassing climate, soil, and historical yield data to create predictive models for legume crops (Na & Na, 2024).

Partial least squares regression (PLSR) has traditionally been employed in hyperspectral remote sensing due to its unique ability to handle collinear and high-dimensional predictor variables (Jin et al., 2022). This is particularly important in the field of remote sensing, where the number of spectral bands can be vast, and the collinearity among them can be overwhelming. While PLSR regression has been the traditional method for hyperspectral remote sensing analysis, the advances in computing power and the growing availability of data have opened new possibilities for analysis through ML techniques such as k-nearest neighbors (k-NN), support vector networks (SVN), and random decision forest (RDF). Incorporating these techniques can lead to new insights and a more complete understanding of the highly managed agricultural environments concerned with livestock feed quality.

When evaluating legume forages for livestock feed, it is important to consider parameters such as neutral detergent fiber (NDF), acid detergent fiber (ADF), in vitro true digestibility (IVTD), and CP. NDF, which measures the cell wall components of forage, is crucial for estimating the bulkiness and filling effect of forage in ruminant diets (Diatta et al., 2021). ADF, a measure of the cellulose and lignin components of forage, is used to calculate forage digestibility (Springer et al., 2023). IVTD is a measure of the extent to which a forage is broken down and digested in the stomach (Varga, G. A., 2023). CP is an approximate measure of total protein content derived from nitrogen concentration. Legume forages typically contain more CP than grasses, serving as a good protein source for livestock (Springer et al., 2023). High-quality legume forages have reduced levels of NDF and ADF, enhanced CP content, and higher IVTD values and they provide better nutrition to livestock (Baath, Baath, et al., 2020; Ball et al., 2001). Moreover, the high values of legume forage qualities provide a greater opportunity for hyperspectral remote sensing to correlate between reflectance and forage quality measures. Utilizing ML techniques to build predictive models from hyperspectral data in legume agriculture represents a tremendous opportunity to improve feed management practices, increase productivity, and promote sustainability (Ang & Seng, 2021).

Overall, numerous studies have demonstrated the ability of hyperspectral reflectance to estimate forage quality traits such as CP, fiber fractions, and digestibility (Ram et al., 2024). Many studies have applied the hyperspectral efforts to various crops such as alfalfa (*Medicago sativa*) quality (Fu et al., 2025), specialty crops (Flynn et al., 2020), and maize (Xue et al., 2024), each which have shown strong predictive performance under controlled or single-species conditions. These studies highlight the promise of hyperspectral analytics for forage quality estimation but also reveal that most work has focused on individual species or uniform canopies under limited management variability.

Despite these advances, few studies have evaluated hyperspectral and forage quality relationships across multiple legume species with contrasting canopy structures, growth habits, and management contexts. Even fewer have examined how these relationships translate to upcoming advancements in satellite-based hyperspectral sensitivities such as CHIME. As a result, the generalizability of hyperspectral

models across diverse legume species and their suitability for upcoming satellite missions remains poorly explored.

Therefore, the objectives of this study are to (1) compare various ML methods (k-NN, PLSR, SVN, and RDF) for developing models to predict NDF, ADF, IVTD, and CP using in situ hyperspectral data; (2) use convolving techniques to match in situ hyperspectral wavelengths to the bands of a hyperspectral satellite (such as CHIME) to compare the performance of different ML methods in predicting NDF, ADF, IVTD, and CP; and (3) to compare CHIME model implications of species inclusion as an input variable versus building models separately by species.

2 | MATERIALS AND METHODS

2.1 | Legume biomass sampling and analytical procedures

The field experiment used three species of legumes, including soybean, tepary bean, and moth bean, as part of a 2-year study (2018 and 2019) carried out at the United States Department of Agriculture–Agricultural Research Service Grazinglands Research Laboratory in El Reno, OK.

Soybean (variety Midland 3926 NR2), moth bean (variety PI 429680), and tepary bean (variety PT0820) were planted on June 14, 2018, and July 3, 2019, using a four-row vacuum planter (Monosem Inc.). Two row spacing treatments, 38 and 76 cm, were implemented for all three species, with intra-row plant spacing maintained at 6 and 3 cm, respectively, to achieve a consistent planting density across experimental plots. Two water regime treatments included irrigated and rainfed conditions, with half of the plots receiving drip irrigation and the remaining half relying solely on natural rainfall. The field experiment was arranged in a split-split plot design, where water regimes were treated as the main plots, species as subplots, and row spacing as sub-subplots. The experimental design comprised four replications of every treatment combination, with each plot measuring 4 m × 3 m. More detailed information on agronomic management practices and experimental treatments can be found in Baath et al. (2024).

Field data collection, which involved both spectral data acquisition and aboveground biomass sampling, was conducted over a span of 2 years, with four sampling events in 2018 and three in 2019. Specifically, data were collected at 40, 48, 53, and 78 days after planting (DAP) in 2018 and 43, 54, and 62 DAP in 2019. During most samplings, soil was visible between rows, and canopy coverage ranged around 50%–70% (Baath et al., 2023). The combination of the field design and sampling dates resulted in 334 spectral and feed quality samples (soybean [$n = 111$], tepary bean [$n = 112$], and moth bean [$n = 111$]). The selection of dates for spectral data collection was based on weather conditions that permitted clear data acquisition without cloud cover at solar noon. Aboveground biomass samples, obtained from 0.5 m row lengths, were subjected to oven-drying at 60°C until a constant weight was achieved and then stored at room temperature for lab analysis. The total N concentration in each sample was determined using the flash combustion method with a Vario Macro model from Elementar Americas, Inc. Subsequently, the concentration was converted to CP by multiplying it by a factor of 6.25. The IVTD was obtained for each sample, following the Daisy Digester procedures from ANKOM Technology. The concentrations of NDF and ADF were determined using batch fiber analyzer techniques.

2.2 | Hyperspectral data acquisition

Hyperspectral data sampling was conducted to acquire in situ reflectance data prior to biomass harvesting. The reflectance data were obtained using a FieldSpec Pro FR spectroradiometer (Malvern Panalytical), which features a spectral resolution of 1.4 nm in the range of 350–1000 nm and 2.0 nm in the range of 1000–2500. To ensure consistency in the spectroradiometer readings, a Spectralon defuse reference panel was utilized as a white reference every 15 min throughout the sampling process. The sensor was set at a height of 1.2 m above the ground while collecting data over the target area, producing a footprint corresponding to 0.5 m for biomass samples taken at the same location. At each site, imaging was performed three consecutive times, with the data averaged to create a unified reflectance curve for each biomass sample.

2.3 | Hyperspectral data processing and band calibration

The Savitzky–Golay filter (15 bands; referred to as SG) (Savitzky & Golay, 1964) was employed to smooth raw spectral curves. This filter eliminates noise, which consists of signal variations that do not correspond to the underlying trend or pattern being analyzed. SG processing is essential for removing noise regions from measured spectra. It is a widely recognized technique for its effectiveness in smoothing and reducing the impact of noise on statistical classification (Gupta et al., 2020). It has been demonstrated to effectively smooth noise while preserving spectral features in low-intensity spectra, making it advantageous for noise reduction compared to other methods (Barton et al., 2021). Additionally, the SG filter is known to be a low-pass filter that works well in noise reduction while preserving peak amplitude information (Alberto et al., 2011).

The first derivative (FD) of each reflectance curve was calculated to emphasize critical sections of the reflectance curve (Han et al., 2023; Yang et al., 2011; Zhou et al., 2021).

Furthermore, spectral regions affected by atmospheric interference, such as those between 1290 and 1495 nm, 1705 and 2045 nm, and 2355 and 2500 nm, along with recognized splicing points within the spectroradiometer at 350–395 nm and 1005–1015 nm, were omitted from the data evaluation (Flynn et al., 2020).

The results we obtained can be applied to satellite platforms in the future, as they will be equipped with sensors that have hyperspectral sensitivities. To extend our results to the satellite level, we generated a third spectral dataset simulating the spectral responses of the CHIME (Nieke & Rast, 2018). This dataset was developed by averaging reflectance data over 10 nm intervals between 400 and 2500 nm, yielding 210 distinct wavebands. Due to the presence of atmospheric interference, we excluded wavebands that contained five or more wavelengths influenced by atmospheric noise and the splicing regions of our sensor (1290–1495 nm, 1705–2045 nm, and 2355–2500 nm). After these exclusions, we retained 148 unique wavebands (referred to as CHIME). Each of the three spectral datasets and their corresponding wavelengths (SG and FD) or convolved wavebands (CHIME) serve as the predictor variables.

2.4 | Performance analytics and measurement

The predictor variables of SG, FD, and CHIME datasets were analyzed alongside laboratory data serving as predicted variables, including NDF, ADF, IVTD, and CP. Various modeling techniques, such as PLSR, k-NN, SVN, and RDF, were employed to analyze the unique combinations of predictor and predicted variables. The first datasets for predictor variables were SG and FD and were handled similarly aside from the spectral post-processing, with FD employing the extra step of calculating the FD of the spectral curve. The models were trained and tested using a 70:30 data split, ensuring a balanced evaluation. To maintain balanced representation across legume types, the 70:30 train–test split was stratified so that each species contributed an equal proportion of samples, supporting more robust and comparable model performance across all legumes. To ensure the modeling's robustness, the K-fold technique was employed 500 times on the training dataset with the testing set being completely withheld from any of the resampling and validation steps. This separation ensures that performance estimates from cross-validation reflect model generalizability without contamination from the test data. The performance metrics presented for each analysis evaluated include mean coefficient of determination (R^2), mean root mean square error (mean RMSE), and component measures (ncomp for PLSR) or their modeling equivalents (such as *k*/*c*/mtry for k-NN, SVN, and RDF, respectively). In the k-NN method, *k* represents the count of closest neighbors used to determine the classification of a new data point. Meanwhile, the *c* parameter in SVN balances the trade-off between maximizing the margin and minimizing the training error, penalizing misclassification. For RDF, mtry denotes the number of features randomly selected at each split of a tree, affecting both the randomness of feature selection and the overall diversity among the trees. For the ML models, we utilized the default hyperparameter settings (see Table S1).

Similar to the SG and FD analytics, next, we conducted the convolving of the bands for the CHIME dataset analytics. Using the same methods described, the analysis results presented here, again, were represented by performance metrics, including the average R^2, the mean RMSE, and a component measure such as ncomp or an equivalent modeling component (e.g., *k*/*c*/mtry).

Incorporating different legume types is essential for enhancing our models' performance, given that we analyzed three distinct legumes with varying growth habits. To ensure the effectiveness of our analysis, we implemented two approaches to account for legume type. To perform the analysis from a pseudo-satellite perspective, we utilized the CHIME dataset and legume type as predictor variables, while the ground-truthed data served as the predicted variables in the first set of model building. In the second set, again, we utilized the CHIME dataset, but this time we split the samples into each of the three legume types to build individual models. The analytics obtained were consistent with those from the previously discussed analyses. Performance metrics were, again, assessed using the average R^2, the average RMSE) and a component measure (e.g., ncomp/*k*/*c*/mtry), depending on the modeling technique. All preprocessing, calibrations, and analytics were performed using the caret packages in the R programming language (R Core Team, 2021; RStudio Team, 2021).

3 | RESULTS

3.1 | NDF, ADF, IVTD, CP, and hyperspectral data

The present study aimed to investigate the forage quality attributes of three legume varieties, namely, soybean, tepary bean, and moth bean. Table 1 presents the results for NDF, ADF, IVTD, and CP for each legume type individually and collectively (all). Findings reveal considerable variations in the forage quality attributes of three legume varieties.

Among the forage quality analytics, soybean was found to have the minimum values of IVTD (55.7%) and CP (7.3%), whereas moth bean and tepary bean exhibited the minimum values of NDF (29.3%) and ADF (20.5%), respectively. Conversely, soybean exhibited the highest values of ADF (54.0%) and CP (30%), while tepary bean (68.7%) and

TABLE 1 Descriptive statistics for soybean, tepary bean, and moth bean (neutral detergent fiber [NDF], acid detergent fiber [ADF], in vitro true digestibility [IVTD], and crude protein [CP]).

Legume(s)	n	NDF(%)				ADF (%)				IVTD (%)				CP (%)			
		Mean	Min.	Max.	Std dev.	Mean	Min.	Max.	Std dev.	Mean	Min.	Max.	Std dev.	Mean	Min.	Max.	Std dev.
Soybean	111	51.2	38.1	65.9	7.8	35.1	22.9	54.0	7.0	73.6	55.7	84.3	6.8	20.1	7.3	30.0	5.4
Tepary bean	112	47.0	33.0	68.7	9.8	33.4	20.5	50.6	6.8	77.1	59.5	85.9	5.4	18.9	7.7	27.2	5.7
Moth bean	111	42.6	29.3	59.1	7.7	29.9	21.1	46.2	4.9	81.1	59.8	90.3	5.3	18.8	9.8	27.2	4.8
All	334	46.9	29.3	68.7	9.1	32.8	20.5	54.0	6.6	77.3	55.7	90.3	6.6	19.3	7.3	30.0	5.3

Abbreviations: Max., maximum; Min., minimum; Std dev, standard deviation.

moth bean (90.3%) exhibited the highest values of NDF and IVTD, respectively. In terms of the mean values, soybean had the highest values of NDF (51.2%), ADF (35.1%), and CP (20.1%), compared to the other legumes. Though moth bean had the highest mean value for IVTD (81.1%). In contrast, moth bean exhibited the lowest mean values for NDF (42.6%), ADF (29.9%), and CP (18.8%). Though, in this case, soybean had the lowest mean value for IVTD (73.6%).

Additionally, the result of the analysis revealed that tepary bean had the highest standard deviations for NDF (9.8%) and CP (5.7%), whereas soybean exhibited the highest standard deviations for ADF (7.0%) and IVTD (6.8%). Conversely, moth bean exhibited the lowest standard deviation for all nutritional attributes studied (Table 1).

The table provides valuable insights into the nutritional composition of legumes. The predicted variables for the other outcomes were similarly represented by these data, gathered from all 112 replicate plots. Furthermore, the hyperspectral triplicate data collection was successful across all plots, apart from one error in two files, leading to sample sizes of $n = 111$ for soybean and moth bean (Table 1). The spectral data acquisition took place before the aboveground collection of legume biomass.

3.2 | Model performance

3.2.1 | In situ hyperspectral

Forage nutrition determination through use of in situ hyperspectral FD data across the four ML models (PLSR, k-NN, SVN, and RDF) is as shown in Table 2.

In general, the SVN model had the highest mean R^2 for all response variables (NDF = 0.731, ADF = 0.609, IVTD = 0.583, and CP = 0.711), while the PLSR model had the lowest mean R^2 (NDF = 0.375, ADF = 0.294, IVTD = 0.236, and CP = 0.469). The NDF response variables had relatively high mean R^2 values across all models except for PLSR (k-NN = 0.610, SVN = 0.731, and RDF = 0.718). The ncomp/*k*/*c*/mtry value varies for each model and response variable, which indicates the specific parameters used for that model. We also explored the effects of the growth stages on the models and found no abnormalities and/or major gaps in the data that would have a substantial influence on the results. Overall, the results suggest that the SVN and RDF models may be the most effective for predicting the response variables analyzed here.

3.2.2 | CHIME

Some similarities and differences were observed between the in situ hyperspectral data and the modified spectral bands

TABLE 2 Machine learning (ML) model results using in situ hyperspectral first derivative (FD) data to model neutral detergent fiber (NDF), acid detergent fiber (ADF), in vitro true digestibility (IVTD), and crude protein (CP). The analytics results are represented by performance metrics, including the average coefficient of determination (R^2), the mean root mean square error (RMSE), and a component measure such as (ncomp) or an equivalent modeling component (e.g., *k*/*c*/mtry).

Model	Predicted	Mean R^2	Mean RMSE	ncomp/*k*/mtry/*c*
PLSR	NDF	0.375	7.266	3
	ADF	0.294	5.575	3
	IVTD	0.236	5.835	3
	CP	0.469	3.900	3
k-NN	NDF	0.610	5.732	5
	ADF	0.502	4.721	5
	IVTD	0.465	4.895	5
	CP	0.604	3.397	5
SVN	NDF	0.731	4.874	1
	ADF	0.609	4.201	1
	IVTD	0.583	4.388	1
	CP	0.711	2.877	1
RDF	NDF	0.718	4.944	54
	ADF	0.575	4.314	54
	IVTD	0.574	4.390	54
	CP	0.710	2.898	54

Abbreviations: k-NN, k-nearest neighbors; PLSR, partial least squares regression; RDF, random decision forest; SVN, support vector networks.

TABLE 3 Machine learning (ML) model results using the convolved first derivative (FD) bands to match CHIME spectral sensitivities data to model neutral detergent fiber (NDF), acid detergent fiber (ADF), in vitro true digestibility (IVTD), and crude protein (CP). Results of the analytics are represented by performance metrics, including the average coefficient of determination (R^2), the mean root mean square error (RMSE), and a component measure such as (ncomp) or an equivalent modeling component (e.g., *k*/*c*/mtry).

Model	Predicted	Mean R^2	Mean RMSE	ncomp/*k*/mtry/*c*
PLSR	NDF	0.409	7.054	3
	ADF	0.341	5.389	3
	IVTD	0.228	5.872	3
	CP	0.426	4.076	3
k-NN	NDF	0.515	6.404	5
	ADF	0.451	4.926	7
	IVTD	0.328	5.479	7
	CP	0.557	3.599	5
SVN	NDF	0.625	5.633	1
	ADF	0.536	4.535	1
	IVTD	0.389	5.227	1
	CP	0.674	3.056	1
RDF	NDF	0.649	5.437	75
	ADF	0.543	4.476	75
	IVTD	0.431	5.011	75
	CP	0.663	3.123	75

Abbreviations: k-NN, k-nearest neighbors; PLSR, partial least squares regression; RDF, random decision forest; SVN, support vector networks.

aligned with the sensitivity profiles of the CHIME satellite. The RDF model performed the best across all variables, with the highest mean R^2 and the lowest mean RMSE for all the forage quality measures except for CP where SVN model resulted in the highest mean R^2 and the lowest mean RMSE (mean $R^2 = 0.674$; mean RMSE = 3.056) (Table 3).

When comparing the findings from the in situ and CHIME datasets, the in situ hyperspectral data had higher mean R^2 and lower mean RMSE values for most models compared to the CHIME in predicting forage quality values. For NDF, the in situ hyperspectral data yielded mean R^2 values ranging from 0.379 to 0.732, while the CHIME data ranged from 0.49

TABLE 4 Machine learning (ML) model results using the convolved first derivative (FD) bands to match CHIME spectral sensitivities data in combination with legume type as a variable to model neutral detergent fiber (NDF), acid detergent fiber (ADF), in vitro true digestibility (IVTD), and crude protein (CP). Results of the analytics are represented by performance metrics, including the average coefficient of determination (R^2), the mean root mean square error (RMSE), and a component measure such as (ncomp) or an equivalent modeling component (e.g., *k*/*c*/mtry).

Model	Predicted	Mean R^2	Mean RMSE	ncomp/*k*/mtry/*c*
PLSR	NDF	0.305	7.591	3
	ADF	0.290	5.565	3
	IVTD	0.365	5.307	3
	CP	0.212	4.742	3
k-NN	NDF	0.599	5.815	5
	ADF	0.502	4.732	5
	IVTD	0.553	4.496	5
	CP	0.506	3.780	5
SVN	NDF	0.723	4.910	1
	ADF	0.607	4.209	1
	IVTD	0.535	4.623	1
	CP	0.686	2.999	1
RDF	NDF	0.731	4.826	76
	ADF	0.553	4.449	76
	IVTD	0.525	4.690	150
	CP	0.665	3.210	76

Abbreviations: k-NN, k-nearest neighbors; PLSR, partial least squares regression; RDF, random decision forest; SVN, support vector networks.

to 0.649. The ADF values showed mean R^2 values between 0.292 and 0.608 and 0.334 and 0.544 for in situ hyperspectral and CHIME data, respectively. The IVTD response variables showed mean R^2 values of 0.234–0.584 and 0.228–0.43 for the in situ hyperspectral and CHIME data, respectively. For CP, the in situ hyperspectral data resulted in mean R^2 values ranging from 0.467 to 0.711, while the CHIME data ranged from 0.426 to 0.674. The highest and lowest percentage variance in R^2 values were observed for IVTD response variables in the SVN model and PLSR model, respectively, with values of 32.87% and 2.56%.

3.2.3 | CHIME and legume type

Incorporating legume type with the CHIME data, the study findings revealed varying performance among different ML models in predicting forage quality. The PLSR model shows the lowest R^2 values (0.212–0.365) across all variables, showing limited ability to explain the variance in the predicted variables. In contrast, the SVN models performed better with higher mean R^2 values and lower mean RMSE values for ADF (mean $R^2 = 0.607$; mean RMSE = 4.209) and CP (mean $R^2 = 0.686$; mean RMSE = 2.999) variables. The RDF model resulted in relatively good performance, particularly for the NDF variable (mean $R^2 = 0.731$; mean RMSE = 4.826), exhibiting high R^2 and relatively low RMSE values (Table 4).

Overall, including the type of legume as a predictor variable in the models resulted in mixed outcomes when assessing forage quality. While the PLSR model showcased a decrease in R^2 values for all forage quality variables with the inclusion of legume type, other models observed an increase in R^2 values for all forage quality variables. Notably, the IVTD exhibited the highest variation among the k-NN, SVN, and RDF models (Figure 1).

The performance of various ML models in predicting the forage nutritional content of legumes with legume type as a predictor variable (LEGUME) as well as modeled individually by type is presented in Figure 2. The results suggest the mean R^2 values for LEGUME are generally lower compared to individual legumes across all response variables and models, except for the k-NN model. In the k-NN model, NDF and IVTD responses were higher for the LEGUME model than for individually modeled legumes by type.

Furthermore, in all models, except for k-NN, the moth bean consistently resulted in higher mean R^2 values compared to other legume types when the performance of different legume types was compared. In the k-NN model, tepary bean in general showed higher mean R^2 values than other legume types.

4 | DISCUSSION

4.1 | Comparative analysis of in situ hyperspectral data and CHIME satellite bands

Hyperspectral remote sensing has great promise to monitor various biophysical and biochemical traits of forage production (Adjorlolo et al., 2015). Research has started

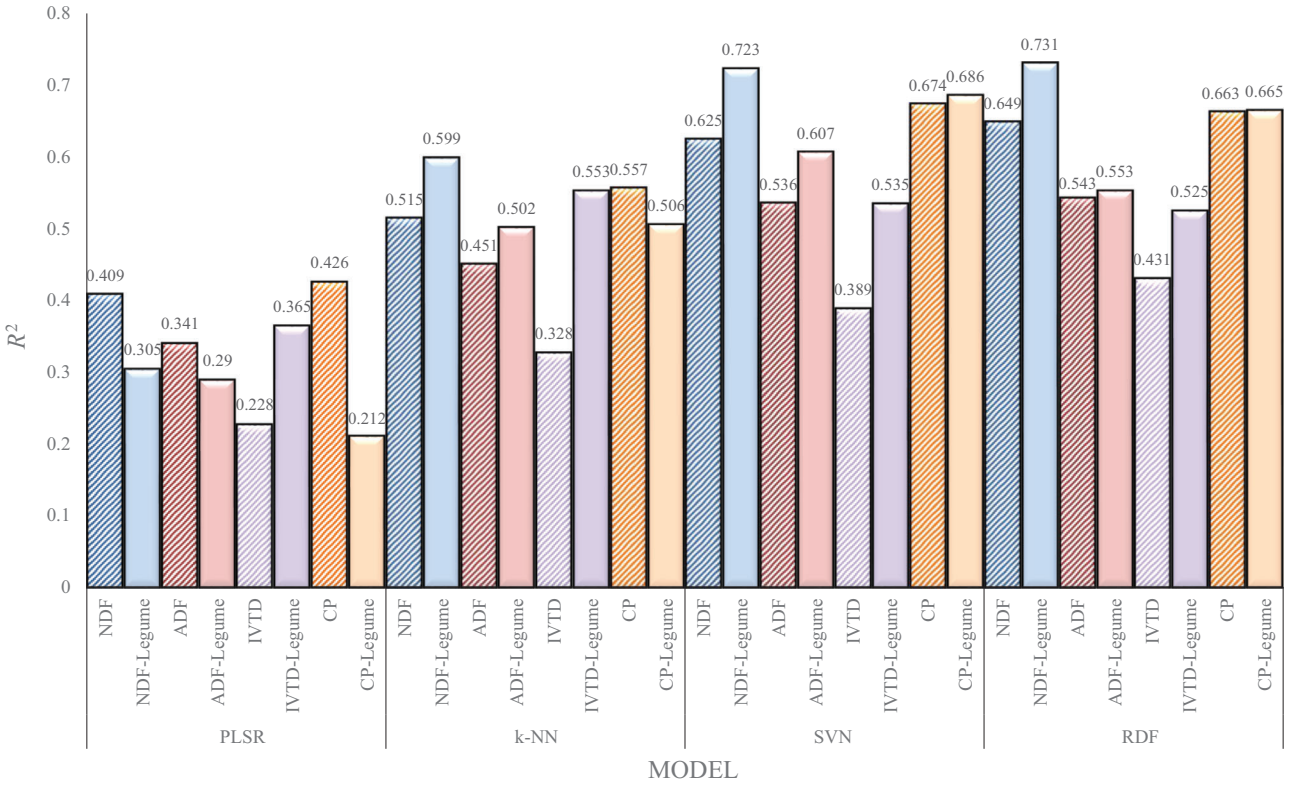


FIGURE 1 Comparison of band-convolved CHIME-only dataset and the CHIME and legume-type dataset measures of performance for each machine learning (ML) model across neutral detergent fiber (NDF), acid detergent fiber (ADF), in vitro true digestibility (IVTD), and crude protein (CP). k-NN, k-nearest neighbors; PLSR, partial least squares regression; RDF, random decision forest; SVN, support vector networks.

to incorporate some hyperspectral satellite sensors to estimate various agricultural characteristics, which could lead to spatial representations through application of precision agriculture and remote sensing methodologies (Dennison et al., 2019; Morata et al., 2023). Our first aim was to analyze the effectiveness of in situ hyperspectral wavelength data and pseudo-satellite convolved wavebands to correspond with the anticipated hyperspectral bands of a future satellite (CHIME) to predict various forage quality measures (NDF, ADF, IVTD, and CP) through application of ML models (PLSR, k-NN, SVN, and CP).

For all hyperspectral analytics, the FD provided the highest mean R^2 values. FD has been found in many studies to outperform SG or raw spectral curves in hyperspectral analytics due to the highlighting of shoulder and transition portions of the reflectance curve (Yang et al., 2011; Zhou et al., 2021). Moreover, as expected, the in situ dataset generally outperformed the CHIME dataset when building models for mostly all quality measures and ML algorithms. This is likely due to greater spectral sensitivities of the in situ data (1503 wavelengths) when compared to the CHIME data (148 wavebands). However, despite the lower spectral resolution of the CHIME dataset, the models for many of the quality measures were still acceptable for prediction purposes (R^2 as high as 0.73). For precision agriculture and management, the pseudo-satellite findings present a great opportunity for the future, as this can lead to improved resource use (i.e., fertilizer, irrigation, etc.) and decision-making (Ahmad & Dar, 2020) when satellites with spectral sensitivities similar to CHIME become available. Additionally, the potential for spatial identification of forage quality from satellite-based remote sensing alongside developing technologies such as virtual fencing (Golinski et al., 2023) provides an opportunity to redirect cattle toward higher quality forages or exclude cattle from areas requiring further forage development (Boyd et al., 2022).

4.2 | ML algorithms for hyperspectral monitoring of forage quality

As computing power and data science technologies continue to evolve, there are now a plethora of ML algorithms available for efficient hyperspectral monitoring in agriculture (Han et al., 2023). In hyperspectral remote sensing, PLSR models are widely and somewhat conventionally utilized in hyperspectral remote sensing (Ang & Seng, 2021; Han et al., 2023; Jin & Wang, 2019). Nevertheless, with the wide range of ML algorithms available, we aimed to compare and identify the most efficient ML algorithms among a commonly used subset (k-NN, PLSR, RDF, and SVN) for the monitoring of forage

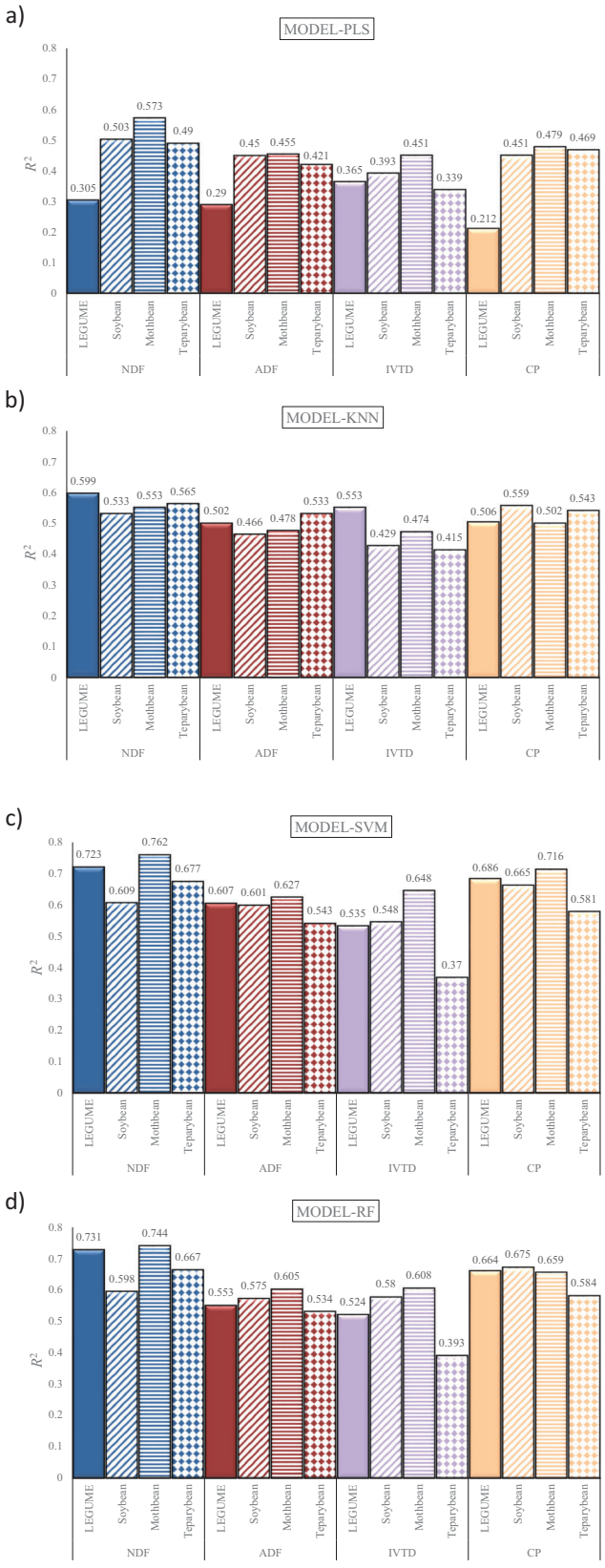


FIGURE 2 Performance comparison of ML models (partial least squares regression: PLSR; k-nearest neighbors: k-NN; support vector networks: SVN; and random decision forest: RDF on band-convolved

(Continues)

FIGURE 2 (Continued)
CHIME and legume type as a predictor variable (LEGUME), and legumes modeled individually by type (soybean, moth bean, and tepary bean), for predicting nutritional parameters (neutral detergent fiber [NDF], acid detergent fiber [ADF], in vitro true digestibility [IVTD], and crude protein [CP]).

quality measures (NDF, ADF, IVTD, and CP) among a set of novel legumes.

k-NN is the preferred choice for predictive studies because of its simplicity, scalability, and comprehensibility. Its ease of use is unmatched when dealing with a diverse range of problems (Li & Wang, 2023). However, among the algorithms tested in the study, k-NN turned out to be the least effective in predicting the suite of variables after PLSR. The RDF and SVM models showed better performance compared to the other models, demonstrating superior predictive performance when evaluating legume characteristics and hyperspectral datasets (in situ and CHIME). Furthermore, our results mostly point to the use of SVN for these analytics over RDF due to its faster computational speed. RDF takes significantly longer to process, roughly 10 times more than other models, which in some instances resulted in >18 h. Incorporating RDF does not offer significant advantages, with only marginal improvements in error reduction (mean RMSE) and mean R^2. In addition, it is crucial to note that the efficiency of this model depends on relative computing power, though if run on the same system, we suspect RDFs would still have greater computational and temporal expense, as seen in other remote sensing studies (Gómez et al., 2016; Noi & Kappas, M., 2018).

Overall, the differences among model performance across forage quality traits likely reflect the degree to which each trait is spectrally expressed. Traits such as CP and NDF have strong localized absorption features in the visual, red-edge, and near-infrared (NIR) portions of the electromagnetic spectrum (Flynn et al., 2020; Liu et al., 2024), enabling more consistent model learning. In contrast, IVTD is an emergent property influenced by multiple biochemical and structural factors, resulting in weaker or more diffuse spectral signatures and sometimes lower predictive performance. Moreover, differences among model performances for the same trait arise from the distinct methods of each algorithm and how it handles high-dimensional spectral data. For instance, PLSR assumes linear relationships and therefore performs best when traits exhibit strong linear spectral responses (Guerri et al., 2024), which was likely not the case here based on the results. In contrast, SVM and RDF can capture nonlinear interactions and are better suited for traits with complex spectral-biochemical relationships (Guerri et al., 2024). The relatively poor performance of k-NN likely reflects the challenges of distance-based methods in high-dimensional hyperspec-

tral spaces where local neighborhoods become sparse and noisy.

4.3 | Legume type considerations

As many agroecosystems aim to incorporate greater diversity to meet various global biodiversity goals (Wanger et al., 2020), identifying remote sensing methods that can accurately assess multiple crop types is paramount. Thus, we aimed to account for the legume types in our remote sensing methods by (1) including the legume type as an additional predictor variable alongside the hyperspectral data (CHIME); and (2) splitting the data into three individual models, one for each legume type. The inclusion of legume type as an additional variable did aid most of the ML models (e.g., k-NN, RDF, and SVN). However, for all quality measures modeled, except IVTD, the inclusion of legume type hindered the PLSR model with a decrease in the R^2 value of up to 0.215. This suggests that when compared to other ML models, PLSR may have greater challenges in remote sensing-based monitoring of biodiverse agroecosystems. Moreover, we wanted to further explore the influence of the legume variable on the ML models. When assessing the top 20 variable ranks, we found that legume type was not among the top contributing variables to the models. The low ranking of legume type in the variable importance analyses suggests that species-level differences were already captured by the spectral data. Because canopy structure and leaf chemistry influence reflectance, the hyperspectral signatures implicitly encode species identity, making the explicit legume-type variable redundant. Additionally, the model findings may also be caused by the introduction of the legume type as a categorical variable as compared to the continuous spectral variables. Models such as PLSR may struggle to work with both categorical and continuous variables simultaneously. Such an effect may not have affected the other models as much as they may have greater flexibility in their model design techniques (Guerri et al., 2024).

Interestingly, in most cases (aside from the relatively low-performing k-NN models for NDF and IVTD), the model building for individual legume types resulted in higher R^2 when compared to the inclusion of legume type as a predictor variable. Moreover, among the individual legume type models, the moth bean models generally resulted in the highest R^2. The moth bean model's greater performance is likely related to the growing pattern of moth bean (prostrate) resulting in greater coverage of the underlying soils that are readily known to cause reflectance noise in many remote sensing applications (Lu et al., 2020). Given the findings, potential application in feed quality assessment seems promising for NDF and CP given their high R^2 values.

5 | CONCLUSION

In this hyperspectral remote sensing study aiming to predict forage quality, we explore the quality of forage from three legume varieties: soybean, tepary bean, and moth bean. Our analysis considers various parameters such as NDF, ADF, IVTD, and CP to assess the quality composition of these legumes. Here, we also delved into predicting forage quality through in situ hyperspectral data versus pseudo-satellite-based data (CHIME) as well as the efficiency of ML models in predicting forage quality. When comparing the in situ hyperspectral data and CHIME data, both were comparable with the former typically producing better modeling results for predicting forage quality. However, the performance of the models varied when incorporating legume type with the CHIME data. The results suggest that satellite-based hyperspectral sensors, such as CHIME, hold the potential for modeling different characteristics of legume forages, especially as additional sensing platforms and data become available. Moreover, from our analysis of various ML models, we found that the SVN and RDF models showed the most favorable results among the ML models examined. Though, employing SVN techniques for processing may prove advantageous, as model building resulted in faster processing times with superior accuracy for models across all characteristics compared to other ML algorithms. It is worth mentioning that the individual models did have a relatively small number ($n = 111$ or 112) of samples, which may affect the reliability/robustness of these models, which provides an opportunity to expand sampling sizes in the future. Nevertheless, the findings within can serve as a guide for future advancements in hyperspectral remote sensing and make valuable contributions to the field of precision agriculture.

AUTHOR CONTRIBUTIONS

Chinmayi H. K.: Conceptualization; formal analysis; investigation; methodology; validation; visualization; writing—original draft. **K. Colton Flynn**: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; supervision; writing—review and editing. **Gurjinder S. Baath**: Formal analysis; methodology; supervision; validation; writing—review and editing. **Trey O. Lee**: Formal analysis; investigation; methodology; validation; writing—review and editing. **Prasanna Gowda**: Investigation; project administration; supervision; writing—review and editing. **Brian Northup**: Formal analysis; investigation; validation; writing—review and editing. **Amanda Ashworth**: Formal analysis; funding acquisition; supervision; validation; writing—review and editing.

ACKNOWLEDGMENTS

This work was supported by an appointment to the Agricultural Research Service administered by the Oak Ridge

Institute for Science and Education. USDA is an equal opportunity provider and employer.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interests.

ORCID

K. Colton Flynn https://orcid.org/0000-0002-5718-1071
Gurjinder S. Baath https://orcid.org/0000-0002-3560-9866
Brian Northup https://orcid.org/0000-0002-7767-4412
Amanda Ashworth https://orcid.org/0000-0002-3218-8939

REFERENCES

Abuova, A., Tulkubayeva, S., Tulayev, Y., Somova, S., & Kizatova, M. (2020). Sustainable development of crop production with elements of precision agriculture in northern Kazakhstan. *Journal of Entrepreneurship and Sustainability Issues*, *7*(4), 3200–3214. https://doi.org/10.9770/jesi.2020.7.4(41)

Adjorlolo, C., Mutanga, O., & Cho, M. A. (2015). Predicting C3 and C4 grass nutrient variability using in situ canopy reflectance and partial least squares regression. *International Journal of Remote Sensing*, *36*(6), 1743–1761. https://doi.org/10.1080/01431161.2015.1024893

Agapiou, A. (2020). Vegetation extraction using visible-bands from openly licensed unmanned aerial vehicle imagery. *Drones*, *4*(2), 27. https://doi.org/10.3390/drones4020027

Ahmad, S. F., & Dar, A. H. (2020). Precision farming for resource use efficiency. In S. Kumar, R. S. Meena, & M. K. Jhariya (Eds.), *Resources use efficiency in agriculture* (pp. 109–135). Springer. https://doi.org/10.1007/978-981-15-6953-1_4

Alberto, D., Rajteri, M., Taralli, E., Lolli, L., Portesi, C., Monticone, E., Jia, Y., Garello, R., & Greco, M. (2011). Optical transition-edge sensors single photon pulse analysis. *IEEE Transactions on Applied Superconductivity*, *21*(3), 285–288. https://doi.org/10.1109/tasc.2010.2087736

Ang, K. L. M., & Seng, J. K. P. (2021). Big data and machine learning with hyperspectral information in agriculture. *IEEE Access*, *9*, 36699–36718. https://doi.org/10.1109/access.2021.3051196

Baath, G. S., Baath, H. K., Gowda, P. H., Thomas, J. P., Northup, B. K., Rao, S. C., & Singh, H. (2020). Predicting forage quality of warm-season legumes by near infrared spectroscopy coupled with machine learning techniques. *Sensors*, *20*(3), 867. https://doi.org/10.3390/s20030867

Baath, G. S., Northup, B. K., Gowda, P. H., Rocateli, A. C., & Singh, H. (2020). Summer forage capabilities of tepary bean and guar in the southern Great Plains. *Agronomy Journal*, *112*, 2879–2890. https://doi.org/10.1002/agj2.20220

Baath, G. S., Northup, B. K., Gowda, P. H., Turner, K. E., & Rocateli, A. C. (2018). Mothbean: A potential summer crop for the Southern Great Plains. *American Journal of Plant Sciences*, *9*, 1391–1402. https://doi.org/10.4236/ajps.2018.97101

Baath, G. S., Sarkar, S., Northup, B. K., Sapkota, B. R., Gowda, P. H., & Flynn, K. C. (2023). Summer pulses as sources of green manure and soil cover in the U.S. Southern Great Plains. *Crop and Environment*, *2*, 66–74. https://doi.org/10.1016/j.crope.2023.04.001

Baath, G. S., Sarkar, S., Sapkota, B. R., Flynn, K. C., Northup, B. K., & Gowda, P. H. (2024). Forage yield and nutritive value of summer legumes as affected by row spacing and harvest timing. *Farming System*, *2*, 100069. https://doi.org/10.1016/j.farsys.2023.100069

Ball, D. M., Collins, M., Lacefield, G. D., Martin, N. P., Mertens, D. A., Olson, K. E., Putnam, D. H., Undersander, D. J., & Wolf, M. W. (2001). *Understanding forage quality* (Publication 1-01). American Farm Bureau Federation.

Barton, S., Alakkari, S., O'Dwyer, K., Ward, T., & Hennelly, B. (2021). Convolution network with custom loss function for the denoising of low SNR Raman spectra. *Sensors*, *21*(14), 4623. https://doi.org/10.3390/s21144623

Boyd, C. S., O'Connor, R., Ranches, J., Bohnert, D. W., Bates, J. D., Johnson, D. D., Davies, K. W., Parker, T., & Doherty, K. E. (2022). Virtual fencing effectively excludes cattle from burned sagebrush steppe. *Rangeland Ecology & Management*, *81*(1), 55–62. https://doi.org/10.1016/j.rama.2022.01.001

Buschkamp, P., Hofmann, J., Rio-Fernandes, D., Haberler, P., Gerstmeier, M., Bartscher, C., Bianchi, S., Delpet, P., Weber, H., Strese, H., & Nieke, J. (2023). CHIME's hyperspectral imager (HSI): Status of instrument design and performance at PDR. In *Proceeding in SPIE* (p. 12777). Society of Photo-Optical Instrumentation Engineers. https://doi.org/10.1117/12.2690341

Chen, S., Zhang, M., Li, X., Li, X., Liu, W., & Ji, B. (2022). Precision agriculture intelligent connection network based on visual navigation. *IET Networks*, *12*(4), 167–178. https://doi.org/10.1049/ntw2.12068

Cho, O. (2024). Machine learning algorithms for early detection of legume crop disease. *Legume Research—An International Journal*, *47*(3), 463–469. https://doi.org/10.18805/lrf-788

Cogliati, S., Sarti, F., Chiarantini, L., Cosi, M., Lorusso, R., Lopinto, E., Miglietta, F., Genesio, L., Guanter, L., Damm, A., Pérez-López, S., Scheffler, D., Tagliabue, G., Panigada, C., Rascher, U., Dowling, T. P. F., Giardino, C., & Colombo, R. (2021). The PRISMA imaging spectroscopy mission: Overview and first performance analysis. *Remote Sensing of Environment*, *262*, 112499. https://doi.org/10.1016/j.rse.2021.112499

Dennison, P. E., Qi, Y., Meerdink, S. K., Kokaly, R. F., Thompson, D. R., Daughtry, C. S. T., Quemada, M., Roberts, D. A., Gader, P. D., Wetherley, E. B., Numata, I., & Roth, K. L. (2019). Comparison of methods for modeling fractional cover using simulated satellite hyperspectral imager spectra. *Remote Sensing*, *11*(18), 2072. https://doi.org/10.3390/rs11182072

Diatta, A. A., Min, D., & Krishna Jagadish, S. V. (2021). Drought stress responses in non-transgenic and transgenic alfalfa—Current status and future research directions. *Advances in Agronomy*, *170*, 35–100.

Finger, R., Swinton, S. M., El Benni, N., & Walter, A. (2019). Precision farming at the nexus of agricultural production and the environment. *Annual Review of Resource Economics*, *11*(1), 313–335. https://doi.org/10.1146/annurev-resource-100518-093929

Flynn, K. C., Baath, G. S., Lee, T. O., Gowda, P. H., & Northup, B. K. (2023). Hyperspectral reflectance and machine learning to monitor legume biomass and nitrogen accumulation. *Computers and Electronics in Agriculture*, *211*, 107991. https://doi.org/10.1016/j.compag.2023.107991

Flynn, K. C., Frazier, A. E., & Admas, S. (2020). Nutrient prediction for tef (*Eragrostis tef*) plant and grain with hyperspectral data and partial least squares regression: Replicating methods and results across environments. *Remote Sensing*, *12*(18), 2867. https://doi.org/10.3390/rs12182867

Fu, S., Liu, J., Gao, J., Feng, Q., Feng, S., Miao, C., Li, Y., Wu, C., & Liang, T. (2025). Improving the estimation accuracy of alfalfa qual-

ity based on UAV hyperspectral imagery by using data enhancement and synergistic band selection strategies. *Computers and Electronics in Agriculture*, *234*, 110305. https://doi.org/10.1016/j.compag.2025.110305

Galvão, L. S., Epiphanio, J. C. N., Breunig, F. M., & Formaggio, A. R. (2018). Crop type discrimination using hyperspectral data. In *Hyperspectral remote sensing of vegetation* (pp. 397–421). CDC Press.

Goliński, P., Sobolewska, P., Stefańska, B., & Golińska, B. (2023). Virtual fencing technology for cattle management in the pasture feeding system—A review. *Agriculture*, *13*(1), 91. https://doi.org/10.3390/agriculture13010091

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008

Gopathoti, K. (2023). Precision agriculture and the role of machine learning. *International Journal of Scientific Methods in Intelligence Engineering Networks*, *01*(06), 01–08. https://doi.org/10.58599/ijsmien.2023.1601

Guerri, M. F., Distante, C., Spagnolo, P., Bougourzi, F., & Taleb-Ahmed, A. (2024). Deep learning techniques for hyperspectral image analysis in agriculture: A review. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, *12*, 100062. https://doi.org/10.1016/j.ophoto.2024.100062

Gupta, S., Huang, C. H., Singh, G. P., Park, B. S., Chua, N.-H., & Ram, R. J. (2020). Portable Raman leaf-clip sensor for rapid detection of plant stress. *Scientific Reports*, *10*(1), Article 20206. https://doi.org/10.1038/s41598-020-76485-5

Guretzky, J. A., Biermacher, J. T., Cook, B., Kering, M. K., & Mosali, J. (2011). Switchgrass for forage and bioenergy: Harvest and nitrogen rate effects on biomass yields and nutrient composition. *Plant and Soil: An International Journal on Plant-Soil Relationships*, *339*, 69–81. https://doi.org/10.1007/s11104-010-0376-4

Han, P., Zhai, Y., Liu, W., Lin, H., An, Q., Zhang, Q., Ding, S., Zhang, D., Pan, Z., & Nie, X. (2023). Dissection of hyperspectral reflectance to estimate photosynthetic characteristics in upland cotton (*Gossypium hirsutum* L.) under different nitrogen fertilizer application based on machine learning algorithms. *Plants*, *12*(3), 455. https://doi.org/10.3390/plants12030455

Hedley, C. (2014). The role of precision agriculture for improved nutrient management on farms. *Journal of the Science of Food and Agriculture*, *95*(1), 12–19. https://doi.org/10.1002/jsfa.6734

Ikhwanti, A., Jayanegara, A., Permana, I. G., Wardani, W. W., Retnani, Y., & Samsudin, A. A. (2020). Sugar, acid soluble polysaccharide, and total phenolic contents in tropical legumes and their relationships with in vitro nutrient fermentability. *Tropical Animal Science Journal*, *43*(4), 331–338. https://doi.org/10.5398/tasj.2020.43.4.331

Ivanovich Vatin, N., Kumar Joshi, S., Acharya, P., Sharma, R., & Rajasekhar, N. (2024). Precision agriculture and sustainable yields: Insights from IoT-driven farming and the precision agriculture test. *Bio Web of Conferences*, *86*, 01091. https://doi.org/10.1051/bioconf/20248601091

Jin, J., & Wang, Q. (2019). Evaluation of informative bands used in different pls regressions for estimating leaf biochemical contents from hyperspectral reflectance. *Remote Sensing*, *11*(2), 197. https://doi.org/10.3390/rs11020197

Jin, J., Wu, M., Song, G., & Wang, Q. (2022). Genetic algorithm captured the informative bands for partial least squares regression better on retrieving leaf nitrogen from hyperspectral reflectance. *Remote Sensing*, *14*, 5204. https://doi.org/10.3390/rs14205204

Jones, S. K., Estrada-Carmona, N., Juventia, S. D., Dulloo, M. E., Laporte, M.-A., Villani, C., & Remans, R. (2021). Agrobiodiversity index scores show agrobiodiversity is underutilized in national food systems. *Nature Food*, *2*, 712–723. https://doi.org/10.1038/s43016-021-00344-3

Kebede, E. (2021). Contribution, utilization, and improvement of legumes-driven biological nitrogen fixation in agricultural systems. *Frontiers in Sustainable Food Systems*, *5*, 767998. https://doi.org/10.3389/fsufs.2021.767998

Kumari, S., Lal, A., Sharma, S., Prikxit, Priyanka, Anjali, Thakur, P., Butail, N. P., & Kumar, P. (2025). A review of multi-dimensional applications of hyperspectral imaging in precision agriculture: Integrating artificial intelligence for scalable solutions. *Remote Sensing Applications: Society and Environment*, *40*, 101808. https://doi.org/10.1016/j.rsase.2025.101808

Ladha, J. K., Peoples, M. B., Reddy, P. M., Biswas, J. C., Bennett, A., Jat, M. L., & Krupnik, T. J. (2022). Biological nitrogen fixation and prospects for ecological intensification in cereal-based cropping systems. *Field Crops Research*, *283*, 108541. https://doi.org/10.1016/j.fcr.2022.108541

Li, W., & Wang, W. C. (2023). Hyperspectral image classification based on spectral feature extraction. *International conference on Image, Signal Processing, and Pattern Recognition* (ISPP 2023) (Vol. *127071B*). Society of Photo-Optical Instrumentation Engineers. https://doi.org/10.1117/12.2681034

Liu, J., Fu, S., Gao, J., Feng, S., Miao, C., Li, Y., Caixia, W., Feng, Q., & Liang, T. (2024). Estimating alfalfa fiber components using machine learning algorithms based on in situ hyperspectral and Sentinel-2 data in the Hexi Corridor region. *Computers and Electronics in Agriculture*, *226*, 109394. https://doi.org/10.1016/j.compag.2024.109394

Lu, B., Dao, P. D., Liu, J., He, Y., & Shang, J. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, *12*(16), 2659. https://doi.org/10.3390/rs12162659

McConnell, M. (2019). Bridging the gap between conservation delivery and economics with precision agriculture. *Wildlife Society Bulletin*, *43*(3), 391–397. https://doi.org/10.1002/wsb.995

Monteleone, S., Moraes, E. A. D., Tondato De Faria, B., Aquino Junior, P. T., Maia, R. F., Neto, A. T., & Toscano, A. (2020). Exploring the adoption of precision agriculture for irrigation in the context of agriculture 4.0: The key role of Internet of Things. *Sensors*, *20*(24), 7091. https://doi.org/10.3390/s20247091

Morata, M., Siegmann, B., Perez-Suay, A., Garcia-Soria, J. L., Rivera-Caicedo, J. P., & Verrelst, J. (2023). Neural network emulation of synthetic hyperspectral Sentinel-2-like imagery with uncertainty. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 762–772. https://doi.org/10.1109/jstars.2022.3231380

Na, M. H., & Na, I. S. (2024). AI-powered predictive modelling of legume crop yields in a changing climate. *Legume Research—An International Journal*, *47*(8), 1390–1395. https://doi.org/10.18805/lrf-790

Nieke, J., & Rast, M. (2018). Towards the Copernicus Hyperspectral Imaging Mission for the Environment (CHIME). In, IGARSS 2018—*2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 157–159). IEEE. https://doi.org/10.1109/IGARSS.2018.8518384

Noi, P. T., & Kappas, M. (2018). Comparison of random forest, k-nearest neighbor, and support vector network classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, *18*(1), 18. https://doi.org/10.3390/s18010018

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ram, B. G., Oduor, P., Igathinathane, C., Howatt, K., & Sun, X. (2024). A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2024.109037

Rao, S. C., Mayeux, H. S., & Northup, B. K. (2005). Performance of forage soybean in the southern Great Plains. *Crop Science*, *45*, 1973–1977. https://doi.org/10.2135/cropsci2004.0598

RStudio Team. (2021). *RStudio: Integrated development environment for R*. RStudio. https://www.rstudio.com/

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*, 1627–1639.

Schimmelpfennig, D. (2017). Crop production costs, profits, and ecosystem stewardship with precision agriculture. *Journal of Agricultural and Applied Economics*, *50*(1), 81–103. https://doi.org/10.1017/aae.2017.23

Schwieder, M., Leitão, P. J., Suess, S., Senf, C., & Hostert, P. (2014). Estimating fractional shrub cover using simulated EnMAP data: A comparison of three machine learning regression techniques. *Remote Sensing*, *6*(4), 3427–3445. https://doi.org/10.3390/rs6043427

Shannon, D. K., Clay, D. E., & Sudduth, K. E. (2018). An introduction to precision agriculture. In D. K. Shannon, D. E. Clay, & N. R. Kitchen (Eds.), *Precision agriculture basics* (pp. 1–12). American Society of Agronomy Crop Science Society of America Soil Science Society of America. https://doi.org/10.2134/precisionagbasics.2016.0084

Singh, H., Northup, B. K., Gowda, P. H., Omara, P., Baath, G. S., & Prasad, P. V. N. (2023). Moth bean and tepary bean as green nitrogen sources in intensive winter wheat cropping systems. *Journal of Agriculture and Food Research*, *15*, 100938. https://doi.org/10.1016/j.jafr.2023.100938

Singh, P., Pandey, P. C., Petropoulos, G. P., Pavlides, A., Srivastava, P. K., Koutsias, N., Deng, K. A. K., & Bao, Y. (2020). Hyperspectral remote sensing in precision agriculture: Present status, challenges, and future trends. In P. C. Pandey, P. K. Srivastava, H. Balzter, B. Bhattacharya, & G. P. Petropoulos (Eds.), *Hyperspectral remote sensing: Theory and applications* (pp. 121–146). https://doi.org/10.1016/B978-0-08-102894-0.00009-7

Springer, R. W., Cherry, N. M., Cross, T. D., Mason, A. C., Guay, K. A., Raub, R. H., Wellmann, K., & Jones, T. N. (2023). 51 estimation of in vitro digestibility measures within equine fecal inoculum using feedstuff fiber composition. *Journal of Equine Veterinary Science*, *124*, 104353.

Syamsi, A. N., Ifani, M., & Subagyo, Y. (2022). The protein-energy synchronization index of the tropical legumes for ruminants. *Journal Peternakan*, *19*(1), 29–37. https://doi.org/10.24014/jupet.v19i1.15415

Talgre, L., Lauringson, E., Roostalu, H., Astover, A., & Makke, A. (2018). Green manure as a nutrient source for succeeding crops. *Plant Soil and Environment*, *58*, 275–281. https://doi.org/10.17221/22/2012-PSE

Tracy, B. F., Maughan, M., Post, N., & Faulkner, D. B. (2010). Integrating annual and perennial warm-season grasses in a temperate grazing system. *Crop Science*, *50*(5), 2171–2177. https://doi.org/10.2135/cropsci2010.02.0110

Varga, G. A. (2023). *In vivo digestibility of forages*. Penn State Extension. https://extension.psu.edu/in-vivo-digestibility-of-forages

Wang, G., Hao, Y., & Chen, Y. (2015). Research on the model of variable-rate fertilization in maize based on geographic information system. In *Proceedings of the 2015 6th international conference on manufacturing science and engineering* (pp. 1756–1763). Atlantis Press. https://doi.org/10.2991/icmse-15.2015.318

Wanger, T. C., Declerck, F., Garibaldi, L. A., Ghazoul, J., Kleijn, D., Klein, A.-M., Kremen, C., Mooney, H., Perfecto, I., Powell, L. L., Settele, J., Solé, M., Tscharntke, T., & Weisser, W. (2020). Integrating agroecological production in a robust post-2020 Global Biodiversity Framework. *Nature Ecology & Evolution*, *4*, 1150–1152. https://doi.org/10.1038/s41559-020-1262-y

Xue, H., Xu, X., & Meng, X. (2024). Estimation model for maize multi-components based on hyperspectral data. *Sensors*, *24*(18), 6111. https://doi.org/10.3390/s24186111

Yang, X., Huang, J., Wu, Y., Wang, J., Wang, P., Wang, X., & Huete, A. (2011). Estimating biophysical parameters of rice with remote sensing data using support vector networks. *Science China Life Sciences*, *54*(3), 272–281. https://doi.org/10.1007/s11427-011-4135-4

Zhou, J., Zhang, Y., Han, Z., Liu, X., Jian, Y., Hu, C., & Dian, Y. (2021). Evaluating the performance of hyperspectral leaf reflectance to detect water stress and estimation of photosynthetic capacities. *Remote Sensing*, *13*(11), 2160. https://doi.org/10.3390/rs13112160

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: H. K., C., Flynn, K. C., Baath, G. S., Lee, T. O., Gowda, P., Northup, B., & Ashworth, A. (2026). Hyperspectral reflectance and machine learning to monitor legume feed quality. *Agronomy Journal*, *118*, e70386. https://doi.org/10.1002/agj2.70386