Portable Visible Hyperspectral Imaging for Prediction of Wheat Grain Protein and Gluten

Gozde Ozdogan
School of Biosystems and Food Engineering
University College Dublin
Dublin, Ireland
gozde.ozdogan@ucd.ie

Aoife Gowen
School of Biosystems and Food Engineering
University College Dublin
Dublin, Ireland
aoife.gowen@ucd.ie

***Abstract*— Traditional wheat quality assessment is destructive and time-consuming. This study evaluates Visible Hyperspectral Imaging (Vis-HSI) under cold (5600K), medium (4400K), and warm (3200K) LED light temperatures to non-destructively predict protein, wet gluten (WG), and dry gluten (DG) in 38 wheat varieties. Gaussian Process Regression (GPR) models were developed using SNV and MSC spectral pre-treatments. Warmer light (3200K and 4400K) significantly enhanced model accuracy. Using Neighborhood Component Analysis (NCA) for feature selection, five key wavelengths were identified, achieving strong prediction performance: WG (R^2_P = 0.92, $RMSE_P$ = 3.49), DG (R^2_P = 0.87, $RMSE_P$ = 1.35), and protein (R^2_P = 0.92, $RMSE_P$ = 0.44). The findings demonstrate that Vis-HSI under warm lighting provides a viable, cost-effective and rapid method for non-destructive wheat quality assessment.**

Keywords—Hyperspectral Imaging, Wheat, Wet Gluten, Protein, Dry Gluten

I. INTRODUCTION

Global consumption of wheat, an ancient and vital crop, rises annually [1]. Its flour uniquely transforms into pliable dough, creating a versatile base for baked goods, including bread, pasta and cakes, which serve as convenient, adaptable staples in today's fast-paced landscape [2].

Among quality parameters such as hardness, moisture, ash and enzyme activity, protein content is the most crucial factor, driving both baking performance and global wheat prices [3]. However, protein quantity alone is insufficient, as it reflects quantity rather than specific composition [1].

Wheat proteins consist of gluten (80%-85%) and non-gluten proteins (15%-20%) fractions [4]. Upon hydration, gluten proteins form a viscoelastic matrix that traps gas and structures baked goods. As this matrix dictates wheat's performance, for instance, high-gluten flour is ideal for bread and pasta, while low-gluten flour suits cakes and cookies, evaluating both total protein and gluten content is essential for a comprehensive quality assessment [5].

Traditional methods for determining protein content include the Kjeldahl and Dumas methods, while gluten content is measured via the glutomatic system [6]. Because these conventional techniques are destructive, slow, chemical-dependent, and unsuitable for whole grains, interest in non-destructive alternatives has grown.

Spectroscopic techniques are prominent non-destructive tools for measuring grain moisture [7], ash [8] and protein [9, 10]. However, they have key limitations: Near-infrared spectroscopy (NIRS) is sensitive to environmental temperature and humidity, while Fourier transform infrared spectroscopy (FT-IR) is highly affected by sample water content. Machine vision systems are another non-destructive option, used to detect insect infestation [11] and classify wheat varieties [12, 13]. However, because they rely solely on visual and spatial surface features, they cannot analyze chemical composition.

In contrast, HSI integrates both spectral and spatial data, making it highly effective for analyzing heterogeneous food products. Consequently, HSI has been used to evaluate wheat hardness, protein, and micronutrients [3, 14, 15].

Despite these advancements, research has not yet applied Vis-HSI to determine protein and wet/dry gluten in whole wheat grains, nor explored how LED color temperature affects prediction performance. Because color temperature (the perceived warmth or coolness of light) can alter visible-range spectra [16]. Accordingly, this study aims to investigate the use of Vis-HSI to assess wheat protein and gluten content from whole wheat grains under three illumination temperatures: cold (5600 K), medium (4400 K), and warm (3200 K).

II. MATERIALS AND METHODS

A. Wheat Samples and Chemicals

To ensure sample diversity, 38 wheat varieties were sourced from producers, breeders, and laboratories across four Turkish regions. After removing impurities, four replicate 10-g samples of each variety were prepared. Potassium dihydrogen phosphate, sodium chloride, and disodium hydrogen phosphate dihydrate (Lennox, Dublin, Ireland; Sigma-Aldrich, Wicklow, Ireland) were obtained for gluten analysis and stored as instructed.

B. Total Protein, Wet Gluten and Dry Gluten Analysis

Total protein and wet gluten contents were determined using the Kjeldahl method and GAFTA procedure, respectively, as detailed in our previous study [17].

C. Hyperspectral Imaging System and Image Acquisition

Hyperspectral images were acquired using a portable hyperspectral imaging camera (Hinalea, 4200, USA) from 400 nm to 750 nm (1.9 nm resolution, 175 bands, 968x608 pixels). The camera was mounted on a tripod (Manfrotto, MK055XPRO3-BHQ2, Ireland) 65 cm above the sample. Illumination was provided by an LED ring light (Venus V29C, Guangdong, China; 16-inch outer, 12-inch inner diameter) at three color temperatures: 3200 K (warm), 4400 K (medium) and 5600 K (cold) (Figure 1). TrueScope Software (v1.7.1, Hinalea Imaging, USA) was used to control the system and to acquire images.

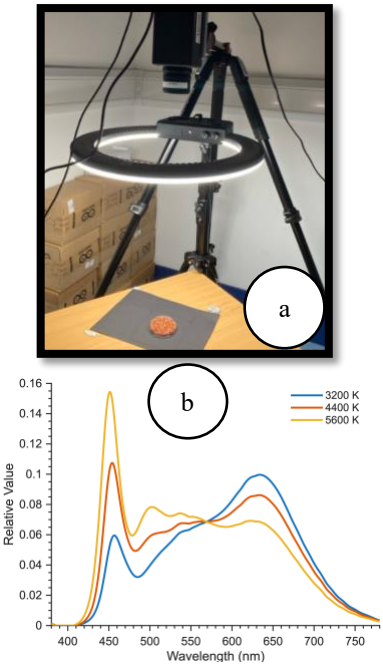


Figure 1. Hyperspectral imaging system: a) Equipment set-up for image acquisition, b) Spectrum of light temperatures (Relative value is the light intensity at each wavelength compared to the peak wavelength).

Following a 1-hour system equilibration, 10-g wheat samples in Petri dishes were imaged on a black background, yielding 152 images per color temperature.

Radiometric correction was performed using a 100% reflectance white reference tile and a lens-capped dark reference, following the equation provided in Reference [18].

D. Data Pre-processing

To remove spectral redundancy and noise, foreground and background pixel spectra were plotted, and the images were cropped to 434–684 nm (126 wavebands). Backgrounds were removed by applying a threshold mask at the highest-contrast wavelength (612 nm), and mean spectra were calculated by averaging the remaining non-zero pixels. Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) pre-treatments were then tested to minimize light-scattering effects caused by wheat shape and surface roughness.

E. Data Exploration and Modelling

Principal Component Analysis (PCA) was used for unsupervised dimensionality reduction, feature extraction, and to explore data variations across different lighting conditions. To predict protein, WG, and DG content, GPR was employed. Unlike PLSR and SVM, which yield a single best-fit function, GPR is a non-parametric Bayesian approach that provides a probability distribution over possible function. Based on 10-fold cross-validation, the GPR models utilized a Matern 5/2 kernel with a sigma of 8.698.

For each wheat variety, one replicate was randomly selected for inclusion in the test set, resulting in 38 spectra, while the remaining three replicates were assigned to the training set, comprising 114 spectra (Data Partition Approach 1). Ten-fold cross-validation was used for hyperparameter optimization. Model performance was assessed using the coefficient of determination and root mean square error for cross-validation (R^2_{CV}, $RMSE_{CV}$) and prediction (R^2_P, $RMSE_P$). All analyses were conducted in MATLAB R2023b (The MathWorks, Inc., Natick, MA, USA).

F. Feature Selection

To improve model efficiency, Neighborhood Component Analysis (NCA) for regression, a supervised, non-parametric feature-weighting technique, was used to select key wavelengths in the Vis range. NCA maximizes leave-one-out validation accuracy using gradient ascent with a regularization term [20] and was applied to standardized predictors to avoid scaling issues. This wavelength selection was performed on the best-performing prediction model.

G. Additional Validation of Selected Features Using Unseen Wheat Varieties (Data Partition Aproach 2)

After identifying the important wavelengths for predicting DG, WG and protein content, an additional validation was performed using a sample division based on wheat variety. In this validation, eight wheat varieties, comprising 32 spectra, were completely withheld as an independent test set, while the remaining 30 varieties, comprising 120 spectra, were used as the training set. Prediction models were then developed using only the five selected wavelengths identified through the feature-selection procedure. The trained models were subsequently evaluated on the spectra from the nine varieties that were not included in model development to assess the generalizability of the selected wavelengths across unseen wheat varieties.

III. RESULTS AND DISCUSSION

A. Spectral Characteristics and Data Distribution

Figure 2 displays the mean spectra and PCA scores. While no distinct peaks appeared in the Vis range (Figures 1a–c), spectral variations primarily reflect differences in pigment concentrations (anthocyanins, carotenoids, and proanthocyanidins) linked to grain color [21]. Divergent trends at 560 nm likely stem from a higher abundance of red kernels compared to yellow ones, while shifting from cold to warm light increased reflectance near 460 nm, which is associated with chlorophyll absorption [22].

The PCA scores plot (Figure 2d) shows distinct clustering between cold and warm light conditions, but not between medium and warm light, confirming that light temperature influences the spectral profiles. Outlying sample scores likely represent red kernel varieties.

Table I. Descriptive statistics for WG, DG and protein were used to develop the calibration and prediction models.

	Subset	Range	Mean	SD	CV	df	F	P
Wet Gluten (%)	**Calibration (n=114)**	22.41-71.20	39.94	12.30	30.80	150	0.00	0.99
	Test (n=38)	22.10-70.85	39.91	12.39	31.04			
Dry Gluten (%)	**Calibration (n=114)**	7.13-23.75	13.63	3.78	27.73	150	0.01	0.92
	Test (n=38)	7.72-23.35	13.55	3.77	27.82			
Protein (%)	**Calibration (n=114)**	9.33-15.50	12.60	1.52	12.06	150	0.00	0.99
	Test (n=38)	9.30-15.55	12.57	1.49	11.85			

Abbreviations: (n= sample number, SD= standard deviation, CV= coefficient of variation (SD/Mean*100), df=degree of freedom, F= Anova coefficient, P= p-value)

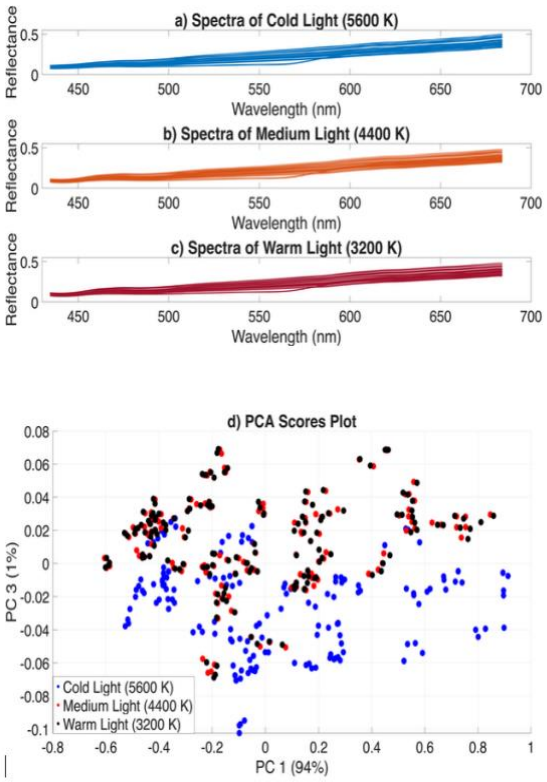


Figure 2. Data exploration plots: (a) Spectra under cold light temperature (5600 K), (b) Spectra under medium light temperature (4400 K), (c) Spectra under warm light temperature (3200 K), (d) PCA scores plot of different illumination adjustments.

To ensure representative model training, reference measurements for the training and test sets were statistically evaluated (Table I). Significant variation was observed in WG (CV = 30.75%), DG (CV = 27.69%), and protein (CV = 12.03%) due to differences in variety and geographical origin. No statistically significant differences were detected between the training and test splits at a 95% confidence level, validating the dataset partitioning.

B. Results of Prediction Models Based on Full Wavelength Range

Table II summarizes the GPR-based predictive performance for WG, DG, and protein content. Warmer illumination (3200 K and 4400 K) outperformed cool light by approximately 18%, with only a slight performance difference between the two warm settings. This improvement may be attributed, at least in part, to the greater representation of longer-wavelength components in warm illumination, which can increase the number of detected photons for a given optical power and consequently improve photon-limited signal-to-noise ratio (SNR).

In addition, warm illumination may enhance contrast and differentiation of wheat grain colour characteristics [16], [21]. The highest coefficients of determination for WG and DG were achieved at 4400 K using SNV-GPR, yielding R^2_P = 0.94 with an $RMSE_P$ = 2.88 and R^2_P = 0.92 with an $RMSE_P$ = 1.09, respectively. Protein content was best predicted at 3200 K using SNV-GPR (R^2_P = 0.93 and $RMSE_P$ = 0.40).

These outcomes outperform the SWIR-PLSR whole-kernel protein predictions in [3], indicating that visible-range spectra and non-linear GPR modelling capture protein-color relationships more effectively. Although flour-based studies achieved high prediction performances, such as [23] (R^2_P = 0.95 for protein via SWIR-PLSR), [24] (R^2_P = 0.86 for protein, R^2_P = 0.85 for WG via RGB-HSI fusion), and [25] (R^2_P = 0.95–0.98 via NIR), our approach yields comparable performance non-destructively using whole kernels in the visible spectrum.

Regarding pre-treatments, MSC and SNV improved warmer-light models by ~10%, correcting for multiplicative scatter caused by kernel roughness and shape [27]. These results demonstrate that visible spectrum colour features are strong proxies for wheat quality traits. Consequently, subsequent wavelength selection was focused on the top-performing 4400 K SNV dataset.

Table II. The results of prediction models based on full wavelength using GPR.

		COLD (5600 K)				MEDIUM (4400 K)				WARM (3200 K)			
		Validation		Test		Validation		Test		Validation		Test	
	Process	R^2_{CV}	RMSE	R^2_P	RMSE	R^2_{CV}	RMSE	R^2_P	RMSE	R^2_{CV}	RMSE	R^2_P	RMSE
Wet Gluten	RAW	0.74	6.23	0.75	6.20	0.66	7.11	0.81	5.39	0.60	7.78	0.82	5.24
	SNV	0.74	6.29	0.76	6.10	**0.94**	**3.04**	**0.95**	**2.88**	0.93	3.24	0.92	3.56
	MSC	0.69	6.86	0.76	6.10	0.94	3.05	0.94	2.89	0.87	4.37	0.92	3.57
Dry Gluten	RAW	0.72	2.01	0.73	1.98	0.59	2.43	0.79	1.75	0.60	2.41	0.85	1.48
	SNV	0.74	1.94	0.76	1.83	**0.92**	**1.06**	**0.92**	**1.09**	0.92	1.10	0.91	1.12
	MSC	0.75	1.87	0.77	1.83	0.91	1.12	0.91	1.10	0.91	1.12	0.91	1.13
Protein	RAW	0.69	0.85	0.80	0.68	0.78	0.71	0.80	0.69	0.76	0.76	0.85	0.58
	SNV	0.76	0.74	0.80	0.69	0.93	0.40	0.92	0.43	**0.92**	**0.42**	**0.93**	**0.40**
	MSC	0.77	0.73	0.80	0.69	0.93	0.41	0.92	0.43	0.92	0.43	0.93	0.41

Abbreviations: (R^2_{CV}= The Coefficient Determination for Cross-validation, $RMSE_{CV}$=Root Mean Squared Error for Cross-Validation, R^2_P = The Coefficient Determination for prediction, $RMSE_P$= Root Mean Squared Error for Prediction)

C. *Results of Prediction Models Based on Selected Wavelengths*

NCA was applied to the SNV-pretreated 4400 K dataset (Table III, Fig. 3). Models were initialised with five selected wavelengths per parameter and systematically reduced to evaluate performance boundaries. All quality parameters shared three wavelengths, 498, 470, and 648 nm, corresponding to chlorophyll and carotenoid pigments [28]. This matches prior findings correlating pigment ratios with protein content [29], [30] and yellow pigment with gluten strength [31].

For WG, the selected bands (498, 470, 620, 522, and 648 nm) yielded $R^2_P = 0.92$, compared to $R^2_P = 0.95$ for the full spectrum. This outperforms flour-based Fourier transform mid-infrared photoacoustic spectroscopy ($R^2_P = 0.89$ using 31 wavelengths [10]) and grain-based NIRS ($R^2_P = 0.88$ using 17 wavelengths [32]). Notably, a three-wavelength model was still at an acceptable performance with an R^2_P of 0.85.

The five-wavelength DG model achieved $R^2_P = 0.87$. Surprisingly, the three-wavelength model achieved slightly higher $R^2_P = 0.90$, indicating that 434 and 684 nm introduced spectral noise. Reference [25] utilized NIRS for predicting the DG of wheat flours, achieving an R^2_P of 0.95. Likewise, [33] used near-infrared transmittance spectroscopy to predict the DG content of whole grains, attaining an R^2_P of 0.96, suggesting that transmittance may offer better accuracy than reflectance. However, our visible reflectance approach is significantly more cost-effective and has comparable results, showing a promising feature of the visible range.

For protein content, the five-wavelength model (470, 498, 648, 434, and 676 nm) achieved $R^2_{CV} = 0.93$ and $R^2_P = 0.92$, with only a 0.04 difference in $RMSE_P$ compared to the full spectrum. This surpasses the single-kernel NIR-HSI R^2_{CV} of 0.79 in [3], though it is lower than the 14-wavelength NIRS performance ($R^2_P = 0.99$) in [34].

These findings confirm that critical wheat quality traits can be resolved using only three to five visible wavelengths, offering a simpler, cost-effective alternative to resource-intensive NIR methods.

Table III. The results of prediction models based on selected wavelengths using GPR.

		Validation Set		Test Set	
	Selected Wavelengths (nm)	**R^2_{CV}**	**$RMSE_{CV}$**	**R^2_P**	**$RMSE_P$**
Wet Gluten	498, 470, 620, 522, 648	**0.95**	**2.66**	**0.92**	**3.49**
	498, 470, 620, 522	0.93	3.24	0.89	4.08
	498, 470, 620,	0.86	4.58	0.85	4.74
Dry Gluten	498, 648, 470, 434, 684	**0.90**	**1.15**	**0.87**	**1.35**
	498, 648, 470, 434	0.89	1.28	0.86	1.43
	498, 648, 470	0.90	1.17	0.90	1.19
Protein	470, 498, 648, 434, 676	**0.93**	**0.40**	**0.92**	**0.44**
	470, 498, 648, 434,	0.93	0.40	0.88	0.54
	470, 498, 648,	0.84	0.61	0.84	0.61

Abbreviations: (R^2_{CV}= The Coefficient Determination for Cross-validation, $RMSE_{CV}$=Root Mean Squared Error for Cross-Validation, R^2_P = The Coefficient Determination for prediction, $RMSE_P$= Root Mean Squared Error for Prediction)

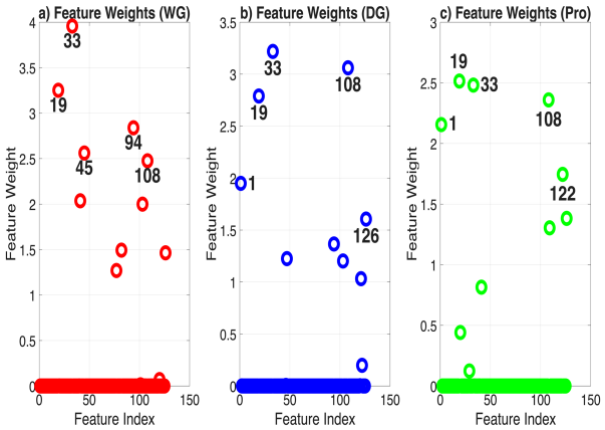


Figure 3.Feature weights plot based on NCA results (a) Feature weight plot for WG prediction, (b) Feature weight plot for DG prediction, (c) Feature weight plot for protein prediction.

D. *Results of Additional Validation of Selected Features Using Unseen Wheat Varieties*

The results for predicting unseen eight wheat varieties' quality parameters are summarized in Table IV. The results with data partition approach 2 were lower than approach 1, however, R^2_P values remained acceptable ranges R^2_P of 0.78 for WG, R^2_P of 0.81 for DG and R^2_P of 0.80 for protein prediction. These findings suggest that WG, DG, and protein content in wheat grains can be predicted with an accuracy exceeding R^2_P of 0.70 using five wavelengths within the visible range, even when applied to previously unknown wheat varieties. This underscores the relevance of the selected spectral features and the significance of the visible range for compositional analysis in the current study. These findings also support the findings in our previous study [35]

Table IV. The results of prediction models based on selected wavelengths using the data partition approach 2

		Validation Set		Test Set	
	Selected Wavelengths (nm)	**R^2_{CV}**	**$RMSE_{CV}$**	**R^2_P**	**$RMSE_P$**
WG	498, 470, 620, 522, 648	0.92	3.68	0.78	4.18
DG	498, 648, 470, 434	0.87	1.46	0.81	1.20
Protein	470, 498, 648, 434, 676	0.92	0.46	0.80	0.50

Abbreviations: (R^2_{CV}= The Coefficient Determination for Cross-validation, $RMSE_{CV}$=Root Mean Squared Error for Cross-Validation, R^2_P = The Coefficient Determination for prediction, $RMSE_P$= Root Mean Squared Error for Prediction)

IV. CONCLUSION

This study successfully demonstrates the feasibility of using Vis-HSI to predict the protein content, WG and DG in whole wheat grains under different light conditions. The results indicate that light temperature significantly influences the prediction performance, with warmer light temperatures (3200K and 4400K) yielding superior results. The use of Vis-HSI under medium light conditions (4400K) proved most effective, especially when combined with spectral pre-

treatments like SNV and MSC, which further enhanced model accuracy.

The study's findings suggest that the visible spectrum contains sufficient information to predict key wheat quality parameters, potentially offering a more practical and cost-effective alternative to traditional methods. The application of NCA for feature selection demonstrated that high prediction accuracy could be achieved with a reduced number of wavelengths, emphasizing the efficiency of Vis-HSI in assessing wheat quality. Notably, three common wavelengths (498 nm, 470 nm, and 648 nm) were identified as particularly informative across different parameters, highlighting the visible spectrum's relevance in evaluating wheat's protein and gluten content.

Using only these five selected wavelengths, the simplified GPR models achieved robust coefficients of determination for prediction: WG (R^2_P = 0.92, $RMSE_P$ = 3.49), DG (R^2_P = 0.87, $RMSE_P$ = 1.35) and protein (R^2_P = 0.92, $RMSE_P$ = 0.44). To further validate the selected wavelengths, the dataset was partitioned according to wheat variety (Approach 2), such that the models were evaluated on wheat varieties not included in the calibration dataset. Using the same five wavelengths, the resulting models demonstrated promising predictive performance for unseen wheat varieties, with R^2_P = 0.78 and an $RMSE_P$ = 4.18 for DG; R^2_P = 0.81 and an $RMSE_P$ = 1.20 for DG; and R^2_P = 0.80 and an $RMSE_P$ = 0.50 for protein content.

These results compare favorably to, and in some cases outperform, more complex and expensive SWIR, NIR, and FTIR-PAS systems that often require flour samples or destructive testing. By demonstrating that key quality traits are strongly linked to visible color characteristics, this research provides a foundation for developing rapid, cost-effective, and non-destructive optical sorting and quality-monitoring systems for the wheat and baking industries.

REFERENCES

[1] J. Ma, B. Zheng, and Y. He, "Applications of a hyperspectral imaging system used to estimate wheat grain protein: a review," *Front. Plant Sci.*, vol. 13, Art. no. 837200, 2022, doi: 10.3389/fpls.2022.837200.

[2] W. S. Veraverbeke and J. A. Delcour, "Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality," *Crit. Rev. in Food Sci. Nutr.*, vol. 42, no. 3, pp. 179-208, 2002, doi: 10.1080/10408690290825510.

[3] N. Caporaso, M. B. Whitworth, and I. D. Fisk, "Protein content prediction in single wheat kernels using hyperspectral imaging," *Food Chem.*, vol. 240, pp. 32-42, 2018, doi: 10.1016/j.foodchem.2017.07.048.

[4] J. R. Biesiekierski, "What is gluten?," *J. Gastroenterol. Hepatol.*, vol. 32, no. S1, pp. 78-81, 2017, doi: 10.1111/jgh.13703.

[5] P. R. Shewry, N. G. Halford, and D. Lafiandra, "Genetics of Wheat Gluten Proteins," in *Adv. Genet.,* vol. 49, J. C. Hall, J. C. Dunlap, and T. Friedmann Eds. Academic Press, 2003, pp. 111-184, doi: 10.1016/s0065-2660(03)01003-4.

[6] S. Zhang, S. Liu, L. Shen, S. Chen, L. He, and A. Liu, "Application of near-infrared spectroscopy for the nondestructive analysis of wheat flour: A review," *Curr. Res. Food Sci.,* vol. 5, pp. 1305-1312, 2022/01/01/ 2022, doi: 10.1016/j.crfs.2022.08.006.

[7] H. Shi and P. Yu, "Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat," *Food Control*, vol. 82, pp. 57-65, 2017, doi: 10.1016/j.foodcont.2017.06.015

[8] P. R. Armstrong et al., "Predicting single kernel and bulk milled rice alkali spreading value and gelatinization temperature class using NIR spectroscopy," *Cereal Chem.*, vol. 99, no. 6, pp. 1234-1245, 2022, doi: 10.1002/cche.10587.

[9] M. H. Ahmad, M. Nache, S. Waffenschmidt, and B. Hitzmann, "A fluorescence spectroscopic approach to predict analytical, rheological and baking parameters of wheat flours using chemometrics," *J. Food Eng.*, vol. 182, pp. 65-71, 2016, doi: 10.1016/j.jfoodeng.2016.03.006.

[10] Q. Liu et al., "Determination of total protein and wet gluten in wheat flour by Fourier transform infrared photoacoustic spectroscopy with multivariate analysis," *J. Food Comp. Anal.*, vol. 106, Art. no. 104349, 2022, doi: 10.1016/j.jfca.2021.104349.

[11] C. Ridgway, E. R. Davies, J. Chambers, D. R. Mason, and M. Bateman, "Rapid machine vision method for the detection of insects and other particulate bio-contaminants of bulk grain in transit," *Biosyst. Eng.*, vol. 83, no. 1, pp. 21–30, 2002, doi: 10.1006/bioe.2002.0096.

[12] S. Majumdar and D. S. Jayas, "Classification of cereal grains using machine vision: I. Morphology models," *Trans. ASAE*, vol. 43, no. 6, pp. 1669-1675, 2000, doi: 10.13031/2013.3107

[13] N. S. Visen, J. Paliwal, D. Jayas, and N. D. G. White, "Image analysis of bulk grain samples using neural networks," *Can. Biosyst. Eng.*, vol. 46, pp. 7.11–7.15, 2004, doi: 10.13031/2013.15002.

[14] X. Jiang et al., "Rapid nondestructive detecting of wheat varieties and mixing ratio by combining hyperspectral imaging and ensemble learning," *Food Control*, vol. 150, p. 109740, 2023, doi: 10.1016/j.foodcont.2023.109740.

[15] W. Zhao et al., "Identification of wheat seed endosperm texture using hyperspectral imaging combined with an ensemble learning model," *J. Food Comp. Anal.*, vol. 121, Art. no. 105398, 2023, doi: 10.1016/j.jfca.2023.105398.

[16] Soft Lights Foundation, "Chapter 11 - Color Temperature.", *Soft Lights Foundation*. [Online]. Available: https://www.softlights.org/chapter-11-color-temperature/. [Accessed Aug. 19, 2024].

[17] G. Özdoğan and A. Gowen, "Unveiling the potential: Harnessing spectral technologies for enhanced protein and gluten content prediction in wheat grains and flour," *Curr. Res. Food Sci.,* vol. 10, Art. no. 101054, 2025, doi: 10.1016/j.crfs.2025.101054.

[18] S. Mahesh, A. Manickavasagan, D. S. Jayas, J. Paliwal, and N. D. G. White, "Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes," *Biosyst. Eng.,* vol. 101, no. 1, pp. 50-57, 2008, doi: 10.1016/j.biosystemseng.2008.05.017.

[19] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.*, vol. 85, pp. 1-16, 2018, doi: 10.1016/j.jmp.2018.03.001.

[20] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, no. 1, pp. 161-168, 2012, doi:10.4304/jcp.7.1.161-168.

[21] J. Lachman, P. Martinek, Z. Kotíková, M. Orsák, and M. Šulc, "Genetics and chemistry of pigments in wheat grain – A review," *J. Cereal Sci.,* vol. 74, pp. 145-154, 2017, doi: 10.1016/j.jcs.2017.02.007.

[22] M. A. Shahin, S. J. Symons, and D. W. Hatcher, "Quantification of mildew damage in soft red winter wheat based on spectral characteristics of bulk samples: A comparison of visible-near-infrared imaging and near-infrared spectroscopy," *Food and Bioprocess Technology*, vol. 7, no. 1, pp. 224-234, 2014, doi: 10.1007/s11947-012-1046-8.

[23] J. Zhang et al., "Rapid determination of protein, starch and moisture content in wheat flour by near-infrared hyperspectral imaging," *J. Food Comp. Anal.*, vol. 117, Art. no. 105134, 2023, doi: 10.1016/j.jfca.2023.105134.

[24] S. Zhang et al., "Estimation of wheat protein content and wet gluten content based on fusion of hyperspectral and RGB sensors using machine learning algorithms," *Food Chem.,* vol. 448, Art. no. 139103, 2024, doi: 10.1016/j.foodchem.2024.139103.

[25] M. Başlar and M. F. Ertugay, "Determination of protein and gluten quality-related parameters of wheat flour using near-infrared reflectance spectroscopy (NIRS)," *Turk. J. Agr. For.*, vol. 35, no. 2, pp. 139-144, 2011, doi: 10.3906/tar-0912-507.

[26] T. Fearn, C. Riccioli, A. Garrido-Varo, and J. E. Guerrero-Ginel, "On the geometry of SNV and MSC," *Chemometrics and Intell. Lab. Sys.*, vol. 96, no. 1, pp. 22-26, 2009, doi: 10.1016/j.chemolab.2008.11.006.

[27] X. He, T. Zhao, F. Shen, Q. Liu, Y. Fang, and Q. Hu, "Online detection of naturally DON contaminated wheat grains from China using Vis-NIR spectroscopy and computer vision," *Biosyst. Eng.*, vol. 201, pp. 1-10, 2021, doi: 10.1016/j.biosystemseng.2020.11.001.

[28] J. Hernandez, G. A. Lobos, I. Matus, A. Del Pozo, P. Silva, and M. Galleguillos, "Using ridge regression models to estimate grain yield from field spectral data in bread wheat (*Triticum aestivum* L.*)* grown under three water regimes," *Remote Sens*., vol. 7, no. 2, pp. 2109-2126, 2015, doi: 10.3390/rs70202109.

[29] Z. J. Wang, J. H. Wang, L. Y. Liu, W. J. Huang, C. J. Zhao, and C. Z. Wang, "Prediction of grain protein content in winter wheat (*Triticum aestivum* L.) using plant pigment ratio (PPR)," *Field Crops Res.*, vol. 90, no. 2, pp. 311-321, 2004, doi: 10.1016/j.fcr.2004.04.004.

[30] X.-l. Jin et al., "Estimation of Grain Protein Content in Winter Wheat by Using Three Methods with Hyperspectral Data," *Int. J. Agric. Bio.*, vol. 16, no. 3, pp/ 498-504, 2014.

[31] F. Hailu and A. Merker, "Variation in gluten strength and yellow pigment in Ethiopian tetraploid wheat germplasm," *Genet. Resour. Crop Evol.*, vol. 55, no. 2, pp. 277-285, 2008, doi: 10.1007/s10722-007-9233-6.

[32] C. Lü et al., "Variable selection based near infrared spectroscopy quantitative and qualitative analysis on wheat wet gluten," in *Proc. SPIE*, vol. 10461, pp. 20-27, 2017, doi: 10.1117/12.2281441.

[33] C. Miralbés, "Prediction chemical composition and alveograph parameters on wheat by near-infrared transmittance spectroscopy," *J. Agr. Food Chem.*, vol. 51, no. 21, pp. 6335-6339, 2003, doi: 10.1021/jf034235g.

[34] D. Ye, L. Sun, B. Zou, Q. Zhang, W. Tan, and W. Che, "Non-destructive prediction of protein content in wheat using NIRS," *Spectrochim. Acta A, Mol. Biomol. Spectrosc.*, vol. 189, pp. 463–472, 2018, doi: 10.1016/j.saa.2017.08.055.

[35] G. Özdoğan and A. Gowen, "Unveiling the potential: Harnessing spectral technologies for enhanced protein and gluten content prediction in wheat grains and flour," *Current Research in Food Science,* vol. 10, p. 101054, 2025/01/01/ 2025, doi: https://doi.org/10.1016/j.crfs.2025.101054.