Near Infrared Technology applications in Agriculture

by

William Yamada

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Biological Systems Engineering

at the
University of Wisconsin-Madison
2025

Date of Final Oral Exam: 05/29/2025
The dissertation is approved by the following members of the Final Oral Committee:
Dr. Matthew Digman, Assistant Professor, Biological Systems Engineering
Dr. Troy Runge, Professor, Biological Systems Engineering
Dr. Zhou Zhang, Associate Professor, Biological Systems Engineering
Dr. João Dorea, Assistant Professor, Animal and Dairy Sciences

Near Infrared Technology applications in Agriculture

William Yamada

Abstract

Precision agriculture demands rapid, non-invasive tools to monitor crop quality and minimize losses, especially during harvest. Traditional laboratory-based analyses, such as wet chemistry and mechanical sieving, are accurate but slow, labor-intensive, and impractical for in-field deployment. This dissertation presents machine learning frameworks that integrate near-infrared (NIR) spectroscopy, self-supervised computer vision, and multimodal data fusion to estimate forage quality and detect pre-harvest losses in real time.

First, handheld NIR spectrometers were evaluated for their ability to estimate nutritional traits in undried forage. Through spectral preprocessing and calibration transfer, the study demonstrated that portable devices can yield accurate predictions of crude protein and digestibility, offering viable alternatives to benchtop systems for on-farm use.

Second, the dissertation addresses grain loss monitoring in row crops by deploying deep learning detectors (YOLO, DETR, YOLOv8) to identify uncollected grain from ground-based RGB-D imagery. These models, trained and validated on wheat, oats, and soybean datasets, demonstrated strong object-level performance and spatial loss quantification, enabling automation of a process that traditionally relies on manual collection and counting.

Third, a late fusion framework was introduced to combine RGB-derived morphometrics with NIR spectral features selected via Partial Least Squares Regression and Variable Importance in Projection (VIP). This approach yielded a Random Forest model achieving a Pearson correlation of $r = 0.968$ and $RMSE_{CV} = 2.16\%$, outperforming previous single-modality methods and enabling practical, edge-compatible deployment.

Finally, RGB images of corn silage were processed using self-supervised Vision Transformers (DINOv2), enabling regression of the Corn Silage Processing Score (CSPS) without manual annotation or kernel segmentation. Performance was further improved using a proxy variable—the proportion of particles above the 4.75 mm sieve (FoG)—which enhanced interpretability and predictive accuracy.

Collectively, this research advances the development of interpretable, deployable, and multimodal sensing systems for agriculture. It demonstrates that computer vision and spectroscopy can be fused to assess silage quality and crop losses accurately, bridging the gap between lab-grade analytics and in-field decision-making for harvest optimization.

Dedication

To my family,
for their unwavering love, patience, and support.

To my parents, Alberto Masashi Yamada and Lindasir Yokoyama Yamada, who taught me the value of persistence.
To my partner, Fabiola Figueroa, whose belief in me never wavered.
And to the many mentors, collaborators, and friends who shared this journey.

Declaration

I hereby declare that this dissertation is my own original work and has not been submitted, in whole or in part, for any degree or professional qualification at any other institution. Any contributions from others and all sources of information have been properly acknowledged through appropriate citations and references.

Signed,

William Yamada

Acknowledgements

Completing this dissertation has been one of the most challenging and rewarding experiences of my life, and it would not have been possible without the support, guidance, and encouragement of many individuals.

First and foremost, I would like to thank my advisor, Dr. Matthew Digman, and co-advisor, Dr. Troy Runge, for their unwavering mentorship, thoughtful guidance, and belief in my work. Their patience and insight helped me grow both as a researcher and as a person. I am also grateful to my committee members for their valuable feedback and support throughout my PhD.

To my colleagues and labmates, thank you for the discussions, collaborations, fieldwork, and many moments of shared frustration and laughter. Your camaraderie made the journey lighter and more meaningful.

To the faculty and staff at the University of Wisconsin–Madison, especially in the Biological Systems Engineering and Computer Science departments, thank you for creating a nurturing academic environment.

To my family, thank you for your unconditional love and support. To my parents, for instilling in me the values of hard work and perseverance. To my partner, for your strength, understanding, and belief in me during every high and low — this accomplishment is as much yours as it is mine.

Finally, I gratefully acknowledge the funding sources and collaborators whose support made this research possible.

Thank you all.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Modern agriculture increasingly relies on rapid, accurate, and non-destructive sensing to improve crop management, reduce input waste, and enhance feed quality. In this context, Near-Infrared (NIR) spectroscopy has emerged as a promising tool for providing real-time chemical analysis across various agricultural workflows—from forage harvest to grain sorting. Its portability, speed, and minimal sample preparation requirements make it especially appealing for in-field and on-machine deployment.

NIR spectroscopy is a form of vibrational spectroscopy that operates within the wavelength range of 750 to 2500 nm, as seen in figure 1.1, which is located just beyond the visible red region in the electromagnetic spectrum. This region is particularly sensitive to molecular overtones and combination vibrations associated with chemical bonds involving hydrogen atoms, such as O–H, N–H, and C–H. These bonds exhibit characteristic absorption bands that can be detected and quantified, enabling the extraction of both qualitative and quantitative information about the chemical and physical composition of a sample.

Although invisible to the human eye, NIR radiation was first discovered in 1800 by Sir Frederick William Herschel, an English astronomer. While measuring temperature changes in various regions of sunlight dispersed through a glass prism, Herschel noticed a distinct thermal effect beyond the red edge of the visible spectrum. He referred to this invisible radiation as “calorific rays,” what we now recognize as near-infrared light [15].

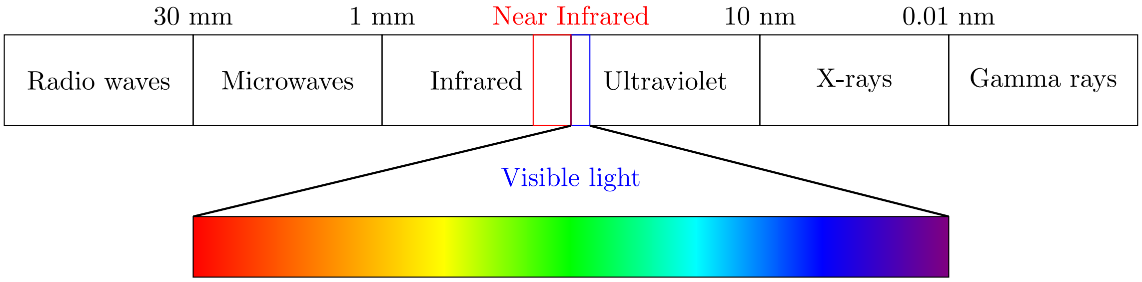


Figure 1.1: The electromagnetic spectrum with the visible light region zoomed in and shown in color.

Despite this early discovery, it took more than a century for NIR to find widespread scientific application. The turning point came in the 1960s, when Karl Norris introduced the use of diffuse reflectance in NIR spectroscopy to evaluate agricultural commodities, particularly for determining the moisture content in cereal grains [28]. Norris's work was groundbreaking: it demonstrated that accurate chemical analysis could be achieved non-destructively, with little to no sample preparation, and directly from heterogeneous, unprocessed samples.

Over subsequent decades, advancements in both instrumentation and computational techniques have significantly expanded the capabilities of NIR spectroscopy. On the hardware side, the development of diode-array detectors, microelectromechanical systems (MEMS), Fabry–Pérot interferometers, and linear variable filters enabled faster acquisition speeds, broader spectral coverage, and greater portability. The miniaturization of spectrometers led to a new generation of field-ready, low-cost NIR devices such as the Viavi MicroNIR, Si-Ware NeoSpectra, and SCiO sensors. These tools allow real-time, in situ measurements, making them suitable for on-farm applications and mobile robotics. NIR devices vary widely in their optical configurations, including dispersive grating systems, Fabry–Pérot interferometers, and MEMS-based filters. Portable devices like the NeoSpectra, MicroNIR, and SCiO typically operate in a reduced wavelength range but maintain enough resolution for applications such as forage and grain quality assessment (Table 1.1).

Simultaneously, the evolution of chemometric tools dramatically improved the inter-

Name	**Parent Company**	**Origin**	**Min λ (nm)**	**Max λ (nm)**
AURORA	GRAINIT	Italy	950	1650
ALBA	GRAINIT	Italy	950	1650
NIR4	AuNIR / AB Vista	England	950	1750
poliSpec	ITPhotonics	Italy	900	1700
SCiO	VeriFood LTD	Israel	740	1070
X-NIR	Dinamica-Generale	Italy	950	1800
EvoNIR 4.0	Dinamica-Generale	Italy	–	–
ASD QualitySpec Trek	Malvern Panalytical	England	350	2500
Enterprise Scanner	Tellspec Inc.	Canada	900	1700
MicroPHAZIR RX Analyzer	Thermo Fisher	USA	1596	2396
CompactSpec - Model EMB	Tec5	Germany	305	1700
FieldSpec Pro	Analytical Spectral Devices	USA	350	2500
NIRQuest 512	Ocean Optics	USA	–	–
Harvest Lab	Carl Zeiss	Germany	950	1530
Labspec2500	Analytical Spectral Devices	USA	350	1830
NIRSystem5000	Foss Electric A/S	Denmark	1100	2498
FoodScan	Foss Electric A/S	Denmark	850	1048
AvaSpec-NIR256-2.5-HSC-EVO	Avantes	Netherlands	1100	2500

Table 1.1: List of common NIR spectrometers found in literature.

pretability and predictive power of NIR spectra. Classical methods like principal component analysis (PCA) and Partial Least Squares Regression (PLSR) have long been the foundation of quantitative NIR calibration [45]. However, more recently, a growing number of studies have leveraged modern machine learning techniques—including support vector machines, random forests, artificial neural networks, and gradient boosting algorithms—to model nonlinear relationships and extract higher-level features from raw spectral data [21]. The integration of unsupervised and semi-supervised learning strategies is also gaining popularity in applications where reference data is limited or expensive to collect.

NIR spectroscopy is thus distinguished by its core advantages: rapid data acquisition, minimal sample preparation, portability, and non-destructive analysis. These attributes define what is now commonly referred to as “*Near-Infrared Technology*”, encompassing not only the spectrometers themselves but also the analytical frameworks, models, and deployment strategies used to interpret and operationalize spectral data in practical settings.

In agriculture, the adoption of NIR spectroscopy has expanded rapidly over the past three decades. Applications span a wide range of domains, including soil nutrient profiling, real-time forage and grain quality assessment, adulteration detection in food products, and crop maturity estimation [5, 8]. On-the-go sensing systems integrated into harvesters and mobile platforms have enabled high-throughput, spatially resolved data collection, laying the groundwork for precision agriculture. The emergence of handheld NIR devices, coupled with open-source chemometric libraries and cloud-based analytics, has further democratized access to this technology among producers, agronomists, and researchers alike.

Despite these advances, the practical deployment of NIR spectroscopy in field conditions continues to face significant challenges. Environmental variability—such as fluctuations in ambient light, temperature, and sample heterogeneity—can reduce model transferability and affect prediction accuracy. Many agricultural decisions also require contextual or structural information that NIR spectra alone cannot provide. For instance, estimating particle size, spatial distribution of biomass, or kernel losses during harvesting requires morphological insight that NIR cannot directly deliver. These limitations have led to growing interest in fusing NIR data with other sensing modalities—such as RGB imaging, depth sensors, or thermal imaging—to provide a more holistic view of plant and product quality.

In this context, the integration of NIR spectroscopy with modern computer vision and machine learning techniques presents a promising avenue for agricultural sensing. By leveraging the complementary strengths of chemical and structural data, and enabling flexible, model-driven workflows, near-infrared technology is positioned to become a cornerstone of smart agriculture and sustainable food systems.

Role of Near-Infrared Spectroscopy in Agriculture

The adoption of Near-Infrared Spectroscopy (NIRS) has accelerated across agricultural sectors due to its speed, non-destructive nature, and ability to capture biochemical information without reagents or intensive sample preparation. NIRS is especially well-suited for in-field analysis, enabling high-throughput, real-time measurements that support precision farming decisions.

Applications in Forage, Grain, and Soil Assessment

In forage quality assessment, NIRS is routinely used to predict crude protein (CP), neutral detergent fiber (NDF), and in-vitro true digestibility (IVTD) — essential indicators for livestock nutrition. Numerous studies have validated its effectiveness on a wide variety of forage types, including legumes, grasses, and silage [42, 29]. NIRS enables producers to monitor feed quality across different harvests and microclimates with high accuracy, often replacing traditional wet chemistry methods that are costly and time-consuming.

In grains, NIRS is employed to estimate moisture, starch, oil, and protein content, facilitating post-harvest sorting, breeding decisions, and quality control [11, 17]. It also supports phenotyping tasks such as identifying drought resistance in corn or characterizing starch quality in sorghum [51, 32].

NIRS is equally impactful in soil science. It enables rapid estimation of organic carbon, total nitrogen, pH, and texture classes—parameters traditionally analyzed using expensive laboratory techniques. Data fusion approaches that combine NIRS with X-Ray Fluorescense (XRF) or RGB imaging have further improved predictive power, particularly in mapping soil fertility and moisture [43, 41]. Real-time, vehicle-mounted NIR systems are being used to generate nutrient prescription maps based on in-situ sensing [44].

Key Advantages in Agricultural Practice

The advantages of NIRS extend across multiple agricultural workflows. Its non-invasive and reagent-free design makes it environmentally safe and operator-friendly. Spectral data can be collected in seconds and processed using pre-calibrated models, allowing near-instant decision-making. Compared to classical methods, NIRS enables decentralized and repeated measurements, supporting dynamic monitoring throughout the growing season.

These strengths make NIRS ideal for use in decentralized environments, such as dairy farms, grain elevators, and autonomous field robots. The versatility of the technique has spurred applications ranging from forage quality control to soil remediation and crop breeding.

Miniaturization and Handheld Devices

Miniaturized NIR devices have expanded the accessibility of this technology to non-specialists. Handheld sensors such as the SCiO, NeoSpectra, and MicroNIR are now being used directly in the field to measure moisture, fiber, and protein [1, 38, 7]. These devices are lightweight, battery-operated, and capable of wireless data transmission. They enable real-time forage quality analysis, pasture monitoring, and silage assessment at the point of need. A few handheld NIR are present in figure 1.2.

However, several limitations remain. Many handheld devices have reduced spectral range (typically 740–1700 nm) and lower spectral resolution compared to benchtop systems [4]. Environmental factors such as ambient light and temperature can also degrade prediction accuracy. Moreover, device-specific calibrations may be required, and model transfer across units often necessitates additional calibration transfer strategies [9]. Calibration transfer methods such as Piecewise Direct Standardization (PDS) [2, 13], External Parameter Orthogonalization (EPO) [40, 25, 23, 50], and Transfer Component Analysis (TCA) [54, 49, 52] are increasingly being explored to align spectral data between instruments. These techniques aim to reduce the need for extensive recalibration by correcting for spectral discrepancies caused by hardware differences, enabling more robust and scalable deployment of predictive models across device types and conditions.

Figure 1.2: Handheld NIR sensors evaluated in this study. From left to right: NeoSpectra (1350–2550 nm, Si-ware Systems Inc., Cairo, Egypt) placed on a turntable; Trinamix (1450–2450 nm, TrinamiX Inc., Ludwigshafen, Germany); and a SCiO Cup (740–1070 nm, Consumer Physics Inc., Tel Aviv, Israel). Photo credit to Dr. Jerry Cherney.

Toward Real-Time, On-the-Go Sensing

The trend toward real-time agricultural sensing is driving the integration of NIR sensors into tractors, harvesters, and robotic platforms. These systems can assess crop or soil properties on the move, enabling continuous quality mapping and adaptive control [46, 30]. On-the-go systems using NIRS are already deployed in self-propelled forage harvesters to estimate moisture content and yield [6]. Additionally, IoT-integrated systems using NIRS have been proposed for pasture management and total mixed ration (TMR) analysis [35].

As machine learning and sensor fusion methods mature, NIR technology is expected to become an even more integral part of digital agriculture. By combining compositional information from NIRS with morphological insights from imaging systems, researchers and producers can derive deeper agronomic insights, optimize input usage, and improve traceability across the food supply chain. In parallel, the emergence of embedded GPU platforms—such as NVIDIA Jetson boards—has enabled real-time, on-device processing of deep learning models. These edge computing solutions are particularly promising for field-deployed systems, allowing for low-latency decision-making and enabling multimodal data fusion directly on robotic or mobile sensing platforms. This development further accelerates the transition toward autonomous, intelligent agricultural diagnostics.

Research Gaps and Challenges

Despite the proven utility of Near-Infrared (NIR) spectroscopy in agricultural applications, several critical challenges limit its broader adoption, particularly in field-deployed, automated decision-making systems. While laboratory studies have shown high prediction accuracy for compositional traits in forages, grains, and soils, these results often fail to generalize to dynamic on-farm environments. Furthermore, many initial studies treated NIR as a standalone solution, neglecting the potential benefits of integrating it with complementary data modalities such as RGB imaging, depth sensing, or spatial information. This section outlines key limitations and identifies underexplored opportunities in the current literature.

Limitations of Single-Modality Sensing

NIR spectroscopy captures the chemical composition of a sample based on molecular overtone and combination vibrations. However, it provides little information about physical structure, morphology, or spatial distribution. This poses a significant limitation in agricultural scenarios where both chemical and physical attributes affect decision-making. For example, particle size distribution in processed forage or grain kernel separation in pre-harvest loss estimation are primarily morphological properties that cannot be inferred from spectra alone.

Additionally, many quality metrics used in agronomic assessments—such as corn silage processing score (CSPS), effective fiber, or lodging severity—are influenced by a combination of structure and chemistry [12]. Relying solely on spectral signatures may thus fail to capture the full variance necessary for robust prediction. In the specific case of CSPS, most existing research relies solely on geometric distributions, kernel size counts, or image-derived particle metrics to approximate processing quality, without integrating chemical composition data [10, 34, 36]. This single-modality approach limits the ability to generalize across environments or machine settings. This gap underscores the need for multimodal approaches that fuse NIR data with visual or spatial features to improve prediction robustness and interpretability.

Challenges in Field Deployment of NIR

The high performance of NIR models in laboratory settings does not always translate to field applications. One primary issue is calibration transferability—models trained under controlled conditions often degrade when applied to heterogeneous samples collected in variable lighting, moisture, or temperature environments [4]. For instance, fresh forage samples collected during harvest may vary significantly in moisture and particle uniformity, leading to spectral shifts that exceed the tolerance of pre-calibrated models.

Spectral variability is another major challenge. Changes in particle orientation, sample compaction, or background reflectance can introduce noise and nonlinearities into the spectra [9, 1]. These sources of variance are difficult to account for using traditional chemometric models, which are often linear or rely on global normalization. Domain adaptation and calibration transfer techniques such as piecewise direct standardization (PDS) [2, 47] and slope/bias correction [48, 22] exist, but their adoption remains limited in portable or embedded systems due to computational constraints.

Additionally, many field applications demand real-time operation, yet most NIR models are still optimized for offline analysis. In on-the-go sensing systems, predictions must be made within milliseconds to support live feedback, which rules out complex or highly tuned preprocessing pipelines.

Opportunities in Multimodal Sensing and Machine Learning Integration

A promising direction to overcome the above limitations lies in multimodal data fusion, particularly the integration of NIR spectroscopy with RGB imaging. RGB data provides rich contextual information about spatial patterns, texture, and color, which can be leveraged to complement the chemical insight offered by NIR. Despite its potential, this approach remains underexplored in the context of forage processing quality and pre-harvest loss detection. Recent work has shown that combining spectral and visual features improves generalizability and robustness under field conditions, but comprehensive studies on multimodal fusion pipelines are still scarce.

Another emerging opportunity is the application of self-supervised learning (SSL) to agricultural sensing [26]. Most NIR calibration models rely on supervised learning with expensive and time-consuming reference data [33]. SSL methods—such as contrastive learning [14, 20, 24], masked reconstruction [31, 16], and cluster-based pretraining [3, 53]—can extract informative representations from unlabeled spectral or image data, reducing dependency on large labeled datasets. These techniques have shown success in other domains, such as medical imaging and material inspection, but their adoption in agricultural NIR workflows remains limited.

Finally, agricultural decision systems increasingly require explainability and interpretability, especially in high-stakes environments like nutrient application or harvest management [39, 19, 27]. Classical linear models like PLSR offer interpretability via regression coefficients and VIP scores, but they often lack flexibility. Deep models, while more expressive, are often treated as black boxes. Research on explainable AI (XAI) techniques for spectral and multimodal data is still nascent in this context [37, 18, 55, 56]. There is a pressing need for models that are not only accurate but also transparent and trustworthy for domain experts.

Together, these gaps in sensor capability, model robustness, and data representation reveal a fragmented landscape for deploying NIR in precision agriculture. This dissertation addresses these challenges through a unified investigation of sensing, learning, and fusion strategies, with the ultimate goal of enabling more reliable, explainable, and field-ready tools for crop quality assessment and harvest decision-making. This gap underscores the need for multimodal approaches that fuse NIR data with visual or spatial features to improve prediction robustness and interpretability.

Table 1.2 summarizes representative studies in this space and highlights the key limitations that motivate the multimodal, self-supervised, and explainable sensing strategies developed in this dissertation.

Study	**Modality**	**Application Domain**	**Limitations / Gaps**
Drewry et al. (2019)	RGB-only	Kernel detection in silage	Uses hand-separated grains for validation; no field-calibrated CSPS reference; no chemical data
Rasmussen et al. (2019)	RGB-only	Forage particle detection	High-performance model (R-FCN) but limited to lab images; no field noise; no NIR or chemical correlation
Rocha et al. (2022)	RGB + ML	Kernel detection in forage	Uses whole kernel percentage as a proxy for CSPS; no spectral features; relies on lab imagery
Digman et al. (2021)	NIR-only	Fresh corn silage quality	NIR device used in situ; strong moisture prediction, but model generalization across seasons not tested
Acosta et al. (2020)	NIR-only	Forage nutritive value	Compares handheld NIR vs. benchtop; does not evaluate multimodal fusion or field deployment
This work	RGB + NIR + SSL	CSPS estimation, forage loss, undried quality	Combines morphology and chemistry; uses self-supervised ViT features and in-field images; interpretable, real-time fusion models

Table 1.2: Summary of representative literature and corresponding research gaps related to silage processing, forage quality, and grain loss sensing.

Research Objectives

This dissertation aims to address critical limitations in current agricultural sensing methods by advancing the application of Near-Infrared (NIR) spectroscopy and integrating it with complementary sensing modalities and machine learning techniques. The specific objectives are:

- **Evaluate the feasibility of using handheld NIR devices for undried forage**

analysis, with a focus on estimating nutritional traits such as crude protein (CP), in-vitro true digestibility (IVTD), and fiber content under real-world sampling conditions.

- **Develop a computer vision pipeline to detect and quantify pre-harvest grain loss in cereal crops**, using RGB imagery and deep learning models trained for fine object detection on in-field images.
- **Integrate RGB-derived morphological features with NIR spectra** to enhance the estimation of forage processing quality, including silage particle distribution and derived metrics such as Corn Silage Processing Score (CSPS).
- **Investigate the use of self-supervised learning techniques for visual feature extraction** from RGB images of silage, and assess their fusion with NIR data for robust and explainable CSPS prediction.
- **Demonstrate late fusion strategies and field deployability for real-time decision support**, leveraging model generalization, multimodal robustness, and explainability as guiding criteria for evaluation.

Organization of the Dissertation

This dissertation is composed of four standalone but thematically connected chapters, each addressing a unique aspect of optical sensing for crop quality monitoring.

Chapter 2 presents a study on the use of handheld NIR spectroscopy to predict undried forage quality. The study evaluates the performance of portable NIR devices under field-relevant moisture and heterogeneity conditions, comparing calibration models against laboratory references.

Chapter 3 explores pre-harvest loss quantification in small grain crops using high-resolution RGB imagery. A deep learning-based object detection framework is proposed and evaluated for its ability to detect and localize grain losses from field robot-acquired images.

Chapter 4 focuses on integrating RGB-derived morphometrics with NIR spectra to estimate forage processing quality. The chapter demonstrates how combining morphological and chemical data improves predictive performance for Corn Silage Processing Score (CSPS) estimation.

Chapter 5 investigates the use of self-supervised learning (SSL) to extract semantic features from silage images. These features are fused with spectral data in a late-fusion framework to enable robust, real-time, and explainable estimation of CSPS.

Chapter 6 concludes the dissertation by summarizing the major findings, discussing their implications for future agricultural sensing systems, and outlining possible directions for future research, including multimodal edge deployment and unsupervised learning methods for calibration transfer.

By building sensing systems that are robust, scalable, and explainable, this work contributes toward the broader goal of developing intelligent agricultural platforms capable of supporting sustainable, data-driven crop and feed management.

References

[1] J. J. Acosta, M. S. Castillo, and G. R. Hodge. “Comparison of benchtop and handheld near-infrared spectroscopy devices to determine forage nutritive value”. In: *Crop Sci.* 60.6 (Dec. 2020), pp. 3410–3422.

[2] M. C. Alamar et al. “Calibration transfer between NIR diode array and FT-NIR spectrophotometers for measuring the soluble solids contents of apple”. en. In: *Postharvest Biol. Technol.* 45.1 (July 2007), pp. 38–45.

[3] H. Alwassel et al. “Self-supervised learning by cross-modal audio-video clustering”. In: *Neural Inf Process Syst* abs/1911.12667 (Nov. 2019). Ed. by H. Larochelle et al., pp. 9758–9770.

[6] D. J. R. Cherney, M. Digman, and J. H. Cherney. “Day-to-day variation in forage and mixed diets in commercial dairy farms in New York”. In: *APPLIED ANIMAL SCIENCE* 37.1 (Feb. 2021), pp. 11–20.

[7] J. H. Cherney, M. F. Digman, and D. J. Cherney. “Handheld NIRS for forage evaluation”. In: *Comput. Electron. Agric.* 190 (Nov. 2021).

[9] M. F. Digman, J. H. Cherney, and D. J. Cherney. “Dry Matter Estimation of Standing Corn with Near-Infrared Reflectance Spectroscopy”. en. In: *Appl. Eng. Agric.* 37.5 (2021), pp. 775–781.

[11] Z. Du et al. “Quantitative assessment of wheat quality using near-infrared spectroscopy: A comprehensive review”. en. In: *Compr. Rev. Food Sci. Food Saf.* 21.3 (2022), pp. 2956–3009.

[13] C. V. Greensill et al. “Calibration transfer between PDA-based NIR spectrometers in the NIR assessment of melon soluble solids content”. en. In: *Appl. Spectrosc.* 55.5 (May 2001), pp. 647–653.

[16] Z. Hou et al. “GraphMAE: Self-supervised masked graph autoencoders”. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, Aug. 2022, pp. 594–604.

[17] M. Huber and K. Jänkälä. “Portable Near-Infrared Spectroscopy for Analysis of Crops”. In: *Cereal Foods World* 65.6 (2020).

[18] R. John Martin et al. “XAI-powered smart agriculture framework for enhancing food productivity and sustainability”. In: *IEEE Access* 12 (2024), pp. 168412–168427.

[19] S. Kawakura et al. “Analyses of diverse agricultural worker data with explainable artificial intelligence: XAI based on SHAP, LIME, and LightGBM”. In: *European Journal of Agriculture and Food Sciences* 4.6 (Nov. 2022), pp. 11–19.

[20] P. Khosla et al. “Supervised Contrastive Learning”. In: *Neural Inf Process Syst* abs/2004.11362 (Apr. 2020). Ed. by H. Larochelle et al., pp. 18661–18673.

[22] L. Li et al. “Calibration transfer between developed portable Vis/NIR devices for detection of soluble solids contents in apple”. In: *Postharvest Biol. Technol.* 183 (Jan. 2022), p. 111720.

[23] J. Liu et al. “Developing a generalized vis-NIR prediction model of soil moisture content using external parameter orthogonalization to reduce the effect of soil type”. en. In: *Geoderma* 419.115877 (Aug. 2022), p. 115877.

[24] X. Liu et al. “Self-supervised learning: Generative or contrastive”. en. In: *IEEE Trans. Knowl. Data Eng.* 35.1 (2021), pp. 1–1.

[25] Y. Liu et al. “Evaluating the characteristics of soil vis-NIR spectra after the removal of moisture effect using external parameter orthogonalization”. en. In: *Geoderma* 376.114568 (Oct. 2020), p. 114568.

[27] P. Naga Srinivasu, M. F. Ijaz, and M. Woźniak. “XAI-driven model for crop recommender system for use in precision agriculture”. en. In: *Comput. Intell.* 40.1 (Feb. 2024).

[29] C. A. Oluk et al. “Determination of Forage Quality by Near-Infrared Reflectance Spectroscopy in Sweet Sorghum (Sorghum bicolor var. saccharatum (L.) Mohlenbr.)” In: *KSU TARIM VE DOGA DERGISI-KSU JOURNAL OF AGRICULTURE AND NATURE* 25.2 (2022), pp. 415–422.

[30] F. Pallottino et al. “Optoelectronic proximal sensing vehicle-mounted technologies in precision agriculture: A review”. en. In: *Comput. Electron. Agric.* 162 (2019), pp. 859–873.

[31] Y. Pang et al. “Masked autoencoders for point cloud self-supervised learning”. In: *arXiv [cs.CV]* (Mar. 2022).

[32] K. H. S. Peiris et al. “Near Infrared Spectroscopic Evaluation of Starch Properties of Diverse Sorghum Populations”. en. In: *Processes (Basel)* 9.11 (Oct. 2021), p. 1942.

[35] G. Rego et al. “A portable IoT NIR spectroscopic system to analyze the quality of dairy farm forage”. en. In: *Comput. Electron. Agric.* 175 (2020), p. 105578.

[37] N. Rodis et al. “Multimodal explainable artificial intelligence: A comprehensive review of methodological advances and future research directions”. In: *IEEE Access* 12 (2024), pp. 159794–159820.

[38] I. R. Rukundo et al. “Comparing the Use of Handheld and Benchtop NIR Spectrometers in Predicting Nutritional Value of Forage”. en. In: *Appl. Eng. Agric.* 37.1 (2021), pp. 171–181.

[39] M. Ryo. “Explainable artificial intelligence and interpretable machine learning for agricultural data analysis”. en. In: *Artificial Intelligence in Agriculture* 6 (2022), pp. 257–265.

[40] F. B. de Santana et al. “Removing the moisture effect in soil organic matter determination using NIR spectroscopy and PLSR with external parameter orthogonalization”. en. In: *Microchem. J.* 145 (Mar. 2019), pp. 1094–1101.

[41] T. R. Tavares et al. “Combined Use of Vis-NIR and XRF Sensors for Tropical Soil Fertility Analysis: Assessing Different Data Fusion Approaches”. en. In: *Sensors (Basel)* 21.1 (Dec. 2020), p. 148.

[42] M. Vranić et al. “Prediction of forage chemical composition by NIR spectroscopy”. en. In: *J. Cent. Eur. Agric.* 21.3 (2020), pp. 554–568.

[43] M. Wan et al. “Estimation of soil pH using PXRF spectrometry and Vis-NIR spectroscopy for rapid environmental risk assessment of soil heavy metals”. en. In: *Process Saf. Environ. Prot.* 132 (2019), pp. 73–81.

[44] W. Wang et al. “Development and performance test of a vehicle-mounted total nitrogen content prediction system based on the fusion of near-infrared spectroscopy and image information”. en. In: *Comput. Electron. Agric.* 192 (2022), p. 106613.

[46] F. Worek and S. Thurner. “11. Yield measurement of wilted forage and silage maize with forage harvesters”. en. In.

[47] F. Wulfert et al. “Correction of temperature-induced spectral variation by continuous piecewise direct standardization”. en. In: *Anal. Chem.* 72.7 (Apr. 2000), pp. 1639–1644.

[48] J. Yang et al. “An interpretable deep learning approach for calibration transfer among multiple near-infrared instruments”. In: *Comput. Electron. Agric.* 192 (Jan. 2022), p. 106584.

[49] J. Yang et al. “Improved calibration transfer between near-Infrared (NIR) spectrometers using canonical correlation analysis”. en. In: *Anal. Lett.* 52.14 (Sept. 2019), pp. 2188–2202.

[50] M. Yang et al. “Removing the moisture effect on predicting soil organic matter using vis-NIR spectroscopy with external parameter orthogonalization”. en. In: *Geoderma Reg.* 37.e00797 (June 2024), e00797.

[51] G. J. Zhang et al. “Comparison of two in situ reference methods to estimate indigestible NDF by near infrared reflectance spectroscopy in alfalfa”. en. In: *Heliyon* 7.6 (June 2021), e07313.

[52] J. Zhang et al. “A partial transfer component regression framework for standard-free calibration transfer of NIR spectroscopic data for soil analysis”. en. In: *Comput. Electron. Agric.* 223.109015 (Aug. 2024), p. 109015.

[53] J. Zhang et al. “Self-supervised convolutional subspace clustering network”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE, June 2019, pp. 5473–5482.

[54] K. Zheng et al. “Calibration transfer of near-infrared spectra for extraction of informative components from spectra with canonical correlation analysis: Using CCA to transfer informative components”. en. In: *J. Chemom.* 28.10 (Oct. 2014), pp. 773–784.

[55] R. Zheng et al. “Optimizing feature selection with gradient boosting machines in PLS regression for predicting moisture and protein in multi-country corn kernels via NIR spectroscopy”. en. In: *Food Chem.* 456.140062 (Oct. 2024), p. 140062.

[56] Q. Zhu et al. “Advanced data-driven interpretable analysis for predicting resistant starch content in rice using NIR spectroscopy”. en. In: *Food Chem.* 486.144311 (Sept. 2025), p. 144311.

Chapter 2

Handheld Near-Infrared Spectroscopy for Undried Forage Quality Estimation

William Yamada, Jerry Cherney, Debbie Cherney, Troy Runge
and Matthew Digman

Publication Note: This chapter is based on the following article published in *Sensors (Basel)*:

Yamada, W., Cherney, J., Cherney, D., Runge, T., & Digman, M. (2024). *Handheld Near-Infrared Spectroscopy for undried forage quality estimation.* **Sensors**, 24(16), 5136. `https://doi.org/10.3390/s24165136`

The author of this dissertation led the data wrangling, spectral modeling, and manuscript preparation.

Abstract: This study investigates the efficacy of handheld Near-Infrared Spectroscopy (NIRS) devices for in-field estimation of forage quality using undried samples. The objective is to assess the precision and accuracy of multiple handheld NIRS instruments—NeoSpectra, TrinamiX, and AgroCares—when evaluating key forage quality metrics such as Crude Protein (CP), Neutral Detergent Fiber (aNDF), Acid Detergent Fiber

(ADF), Acid Detergent Lignin (ADL), in vitro Total Digestibility (IVTD)and Neutral Detergent Fiber Digestibility (NDFD). Samples were collected from silage bunkers across 111 farms in New York State and scanned using different methods (static, moving, and turntable). The results demonstrate that dynamic scanning patterns (moving and turntable) enhance the predictive accuracy of the models compared to static scans. Fiber constituents (ADF, aNDF) and Crude Protein (CP) show higher robustness and minimal impact from water interference, maintaining similar R^2 values as dried samples. Conversely, IVTD, NDFD, and ADL are adversely affected by water content, resulting in lower R^2 values. This study underscores the importance of understanding the water effects on undried forage, as water's high absorption bands at 1400 and 1900 nm introduce significant spectral interference. Further investigation into the PLSR loading factors is necessary to mitigate these effects. The findings suggest that, while handheld NIRS devices hold promise for rapid, on-site forage quality assessment, careful consideration of scanning methodology is crucial for accurate prediction models. This research contributes valuable insights for optimizing the use of portable NIRS technology in forage analysis, enhancing feed utilization efficiency, and supporting sustainable dairy farming practices.

2.1 Introduction

Understanding and managing the nutritional variability of forages is crucial for optimizing dairy farm management and enhancing animal health. The nutritive value of forages, including alfalfa-grass haylage and corn silage, exhibits considerable variability, which can influence milk production efficiency and environmental sustainability [12]. Recent advancements have leveraged handheld near-infrared (NIR) spectroscopy devices, such as the NeoSpectra (1350–2550 nm, Si-ware Systems Inc., Cairo, Egypt), TrinamiX (1450–2450 nm, TrinamiX Inc., Ludwigshafen, Germany), and Agrocares Scanners F-Series (1300–2550 nm, AgroCares, Wageningen, the Netherlands). These tools offer rapid on-site prediction of forage quality, enabling nutritional management by detecting variations in forage nutritive values, including dry matter (DM) [19, 8, 10, 9], Crude Protein (CP) [1, 19, 8, 10, 11, 6, 9, 2, 20], and fiber contents and properties, such as actual Neutral Detergent Fiber (aNDF) [1, 19, 11, 6, 2, 20], Neutral Detergent Fiber Digestibility (NDFD) [10, 11, 6], Acid Detergent Fiber (ADF) [1, 19, 8, 10, 11, 6, 2, 20], Acid Detergent Lignin (ADL) [19, 10, 11, 6], and in vitro Total Digestibility (IVTD) [1, 10, 11, 6, 20]. By facilitating daily adjustments to animal diets based on accurate, real-time forage analysis, handheld NIR devices can significantly enhance the efficiency of feed utilization, reduce environmental impact, and improve the overall profitability and sustainability of dairy farming operations [8, 4, 17].

The interest in compact, handheld spectrometers for analyses conducted directly on farms is gaining momentum [3, 19, 25]. These NIR devices have proven to have good performance on pre-treated samples of dried and ground forage [1]. These portable spectrometers are becoming more common on machinery used for forage harvesting and have also been modified for use with other agricultural implements like liquid manure spreaders [5, 7]. In a study [13], it was found that a portable device (HarvestLab$^{\text{TM}}$ 3000, Deere & Company, Moline, IL, USA) could approximate the quality of a mix of undried forage species, albeit with systematic errors that could be adjusted for accuracy. An essential element for the effectiveness of NIR technology in these applications is the development of a reliable calibration.

The utilization of handheld Near-Infrared Spectroscopy (NIRS) devices in forage assessment aims to permit direct sample analysis without the need for prior sample processing. This is enabled by calibrations developed for wet, unprocessed forage. However, deploying these devices faces several obstacles, notably the influence of the moisture absorption band and sample heterogeneity [16, 21]. Additional operational challenges of handheld NIRS technology include managing the signal-to-noise ratio, ensuring a consistent power supply in a portable format, and maintaining functionality amid diverse and potentially adverse environmental conditions [8, 11].

These studies underscore the practical considerations and performance of handheld

NIRS instruments in the field. For instance, [11] describes the development of predictive models for various forage types, highlighting the significant influence of the scanning methodology on the accuracy of the constituent prediction. Similarly, [8] provides an assessment of multiple handheld NIR devices, examining their precision and accuracy in on-farm forage evaluation, with a focus on dry matter content compared to traditional moisture meters and the robustness of available calibrations for nutritive value determination.

The ongoing development of NIRS technology for forage quality prediction indicates that miniaturized instruments have similar predictive power as benchtop instruments. As this technology is increasing in popularity, it is essential to understand if different portable instruments and scanning patterns affect the quality of the prediction. Thus, the objectives of this research are as follows:

- To assess the precision and accuracy of multiple handheld Near-Infrared Spectroscopy (NIRS) devices when used for on-farm forage evaluation, particularly focusing on the robustness of calibrations for nutritive value determination;
- To examine if different portable instruments and scanning patterns influence the quality of prediction;
- To evaluate the effects of using dried unground samples for forage quality prediction.

2.2 Materials and Method

2.2.1 Samples and Reference Analysis

Predictive NIRS models were developed using NIRS spectra and laboratory reference values for 600 silage samples of mixed haylage. Silage samples were collected between 2021 and 2023 from silage bunkers on 111 farms around New York State. After collection, the samples were vacuum-packed in oxygen-limiting polyethylene bags using a commercial vacuum packing machine for scanning at a later date.

The acquisition of NIR spectroscopic measurement data was achieved using three scanners: NeoSpectra (1350–2550 nm, Si-ware Systems Inc., Cairo, Egypt), TrinamiX

(1450–2450 nm, TrinamiX Inc., Ludwigshafen, Germany), and AgroCares (1300–2550 nm, AgroCares, Wageningen, the Netherlands). The data collected with TrinamiX and AgroCares reported spectra from 1454 to 2446 nm at a fixed step of 4 nm, while the NeoSpectra scanner reported spectra from 1350 to 2550 nm at a variable step between 2.5 and 8.8 nm and a wavelength resolution of 16 nm. Each scanner used different detector types as shown in Table 2.1.

Table 2.1: Instruments and their characteristics.

Property	AgroCares F-series	TrinamiX	NEO Spectra
Detector Type	MEMS	Linear Variable Filter	MEMS-FT-NIR
Spectral Range (nm)	1,450–2,450	1,450–2,450	1,350–2,500
Sample Scanning	Contact	Contact	Contact

Before the scanning process began, all samples were thoroughly mixed in a large plastic container to ensure homogeneity. All samples were scanned in a controlled laboratory environment to avoid any interference of humidity or temperature on the scans. Two primary methods were employed to capture spectra, alongside a third specialized technique. The first method involved placing the scanner's lens in direct contact with the sample, where it remained stationary. The second method required the scanner to be moved across the sample surface during the scanning period, maintaining continuous contact. After each scan conducted with the second method, the samples were mixed again to ensure consistency; this method was tested with the AgroCares and NEOSpectra scanners. The third technique utilized a rotating dish accessory (Si-ware Systems Inc., Cairo, Egypt) for the NEOSpectra instrument, allowing the sample to be scanned continuously. These methods were sequentially applied to each forage sample, with five replicate scans being collected to ensure accuracy and repeatability.

Post-acquisition, the samples were desiccated using forced-air ovens until a consistent mass was achieved at 60 °C, subsequently ground with a Wiley mill (Thomas Scientific, Swedesboro, NJ, USA) through a 1 mm mesh screen, and then stored in plastic containers. The forage constituents appraised included Neutral Detergent Fiber (aNDF), in vitro Total

Digestibility (IVTD), Neutral Detergent Fiber Digestibility (NDFD), Acid Detergent Lignin (ADL), Acid Detergent Fiber (ADF), ash, and Crude Protein (CP), which served as reference variables for the calibration of Near-Infrared Spectroscopy (NIRS) predictive models.

2.2.2 Wet Chemistry

For the chemical analysis, the methodologies aligned with those delineated in the literature [22]. Concisely, forage samples were apportioned into ANKOM F57 filter bags (ANKOM Technology, Macedon, New York, NY, USA) to quantify NDF, ADF, ADL, and 48-hour IVTD. To alleviate gaseous pressure, the filter bags were intermittently removed from their respective containers on both the initial and subsequent days. The digestibility of the Neutral Detergent Fiber was quantified in terms of the percentage of fiber hydrolyzed, with the values expressed on an NDF basis.

The nitrogen (N) content was measured through a combustion technique using a LECO CN628 analyzer (DairyOne, Ithaca, NY, USA), with Crude Protein (CP) being inferred from nitrogen values using the conversion factor of 6.25 as per AOAC guidelines (1995). Duplicate analyses were performed for all constituents, with nitrogen content being quantified in duplicate on a select sample subset to establish the Standard Error for the CP measurement. The laboratory's Standard Error (SEL) pertinent to these determinations has been documented in prior studies that used the same samples but with different instruments and scanning analysis [8, 11].

2.2.3 Model Calibration

To ensure methodological consistency and mitigate the risk of overfitting, uniform data preprocessing and training protocols were employed across all instrument models. Spectral data from the NEOSpectra device were interpolated to achieve a consistent interval of 4 nm. All spectral data were converted to absorbance by employing the logarithmic transformation of the reciprocal reflectance, denoted as log(1/R). Data preprocessing was standardized using a Savitzky–Golay filter with a window length of four, a polynomial order of three, and a derivation order of one. The algorithm of choice for the modeling was Partial Least Squares Regression (PLSR), using Python 3.10.12 and the packages scipy (to preprocess the data) and scikit-learn (to calibrate the PLSR model). In order to have a baseline model for each predicted variable, we opted to work with PLS-1 to understand how the scans affected the individual performance of the models. The selection of the optimal number of latent variables within the range of 1 to 20 was systematically determined using a grid search.

The dataset was divided into 90%/10% for training with five-fold cross-validation (CV) and a separate validation dataset, respectively. We randomly selected bunkers to split the dataset, ensuring that the training and test sets were independent. This approach uses 540 samples for calibration (432 for training and 108 for CV) and 60 samples for testing the final model, providing a robust evaluation of the model's generalizability. By having an external dataset for validation, created as described, we can effectively verify the overfitting of our model [18]. Overfitting can be identified by comparing performance metrics between the training and validation datasets. If the model performs significantly better on the training data than on the validation data or the validation performance deteriorates while the calibration improves, it is likely overfitting. In our study, we adopt a robust outlier detection method utilizing Partial Least Squares (PLS) regression tailored for Near-Infrared (NIR) spectroscopy data analysis. This approach leverages Q-residuals and Hotelling's T-squared statistics to identify deviations, ensuring outliers that could skew the model's predictive accuracy are effectively recognized, using a 95% confidence level [15]. This technique provides a systematic way to refine datasets for better analytical outcomes.

2.2.4 Evaluation

The calibration models were evaluated using the standards set forth by Malley et al. [14] and Williams et al. [24] as shown in Table 2.2. This section details the predictive performance of the models using a suite of metrics, including root mean square error ($RMSE$), Bias, Standard Error (SE), Cross-Validation Standard Error ($SECV$), Coefficient of Determination (R^2), Cross-Validated R^2 (R^2_{CV}), Ratio of Performance to Deviation (RPD), Cross-Validated RPD (RPD_{CV}), and the number of latent variables (LVs) employed on the PLS calibration.

Table 2.2: Calibration classification and level of success according to [14, 24].

Level of Success [14]	R^2 [14]	RPD value [24]	Classification [24]	Application [24]
Not useful	<0.80	<2.0	Very poor	Not recommended
Moderately Successful	0.80 to 0.90	2.0 to 2.5	Poor	Rough screening
		2.5 to 3.0	Fair	Screening
Successful	0.90 to 0.95	3.0 to 3.5	Good	Quality control
		3.5 to 4.0	Very good	Process control
Excelent	>0.95	>4.0	Excellent	Any application

RPD - Ratio of Perfomance to Deviation

We will compare our results with those obtained from various handheld NIRS devices reported in the literature. Specifically, we will reference studies that utilized different instruments on both dried and undried materials. For undried and unground material, we will compare our findings with those from the Aurora device as reported by Cherney et al. (2021) [8] and the NEOSpectra device as reported by Feng et al. (2023) [11]. For dried and ground material, comparisons will be made with results obtained using the MicroPHAZIR (1600–2400 nm, ThermoFisher Scientific, Waltham, MA, USA) and DLP NIRscan Nano EVM (900–1700 nm, Texas Instruments, Dallas, TX, USA) devices as described by Acosta et al. (2020) [1], the NEOSpectra device as reported by Digman et al. (2022) [10], the ASD QualitySpec (350–2500 nm, Malvern Panalytical, Cambridge, UK) and Tellspec (900–1700 nm, Tellspec Inc., Toronto, ON, Canada) devices as detailed by Rukundo et al. (2021) [19], and the Aurora (950–1650 nm, GraiNit S.r.l., Padua, Italy), NIR-S-G1 (950–1650 nm, InnoSpectra, Hsinchu, Taiwan), and SCiO (740–1070 nm, Consumer Physics, Hod Hasharon, Israel) devices as discussed by Berzaghi et al. (2021) [6]. This comprehensive comparison will provide a robust context for evaluating the performance and accuracy of our results and evaluating the impact of using undried materials. In addition, to evaluate the effects of water absorption, we will analyze the major PLS loading components in relation to the water absorption bands as described by Williams [23]. The primary water absorption bands are detailed in Table 2.3.

Table 2.3: Positions of main absorption bands in water.

Wavelength (nm)	1460	1778	1904	2208	2384
Relative Intensity	Large	Very Small	Very Large	Very Small	Very Small

2.3 Results and Discussion

2.3.1 Spectral data

The average spectra and range of the 600 samples scanned are shown in figure 2.1, which presents the mean spectral signatures captured by each instrument. In accordance with the observations reported by Feng et al. [11], overtone bands attributable to O-H bonds are discernible at approximately 1,400 and 1,900 nm, which is consistent with the presence of moisture in the un-dried forage samples. The statistical information of the laboratory measurements of the constituents is provided in table 2.4.

Table 2.4: Laboratory reference values statistics.

	IVTD	aNDF	NDFD	ADF	ADL	CP
Unit			%DM			
Count			600			
Mean	79.22	50.13	58.84	37.18	7.35	17.62
SD	7.31	10.42	9.39	5.80	2.18	4.43
Min	38.13	28.81	11.40	24.22	3.12	6.12
Median	80.73	48.75	58.50	36.50	7.06	18.04
Max	92.92	81.60	80.87	59.06	20.60	27.72

IVTD - in-vitro Total Digestibility, aNDF - actual Neutral Detergent fiber, NDFD - Neutral Detergent Fiber Digestibility, ADF - Acid Detergent Fiber, ADL - Acid Detergent Lignin, CP - Crude Protein, DM - Dry Matter, SD - Standard Deviation

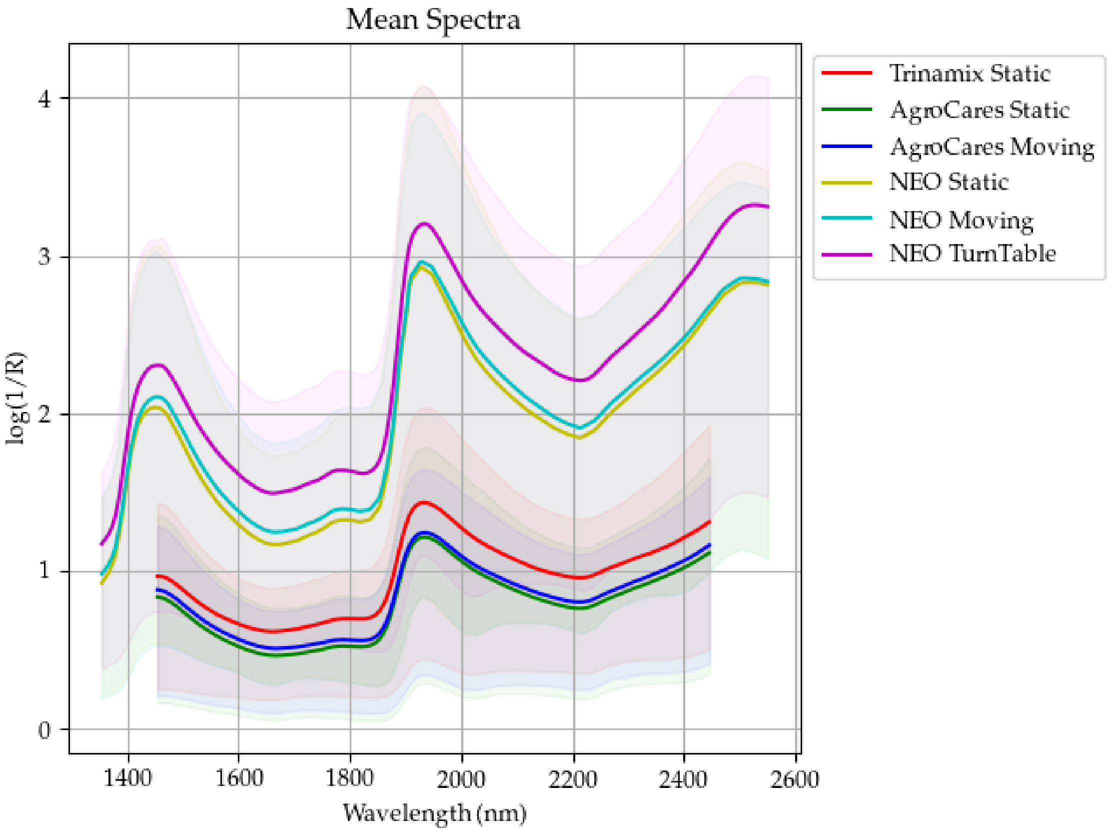


Figure 2.1: Comparative Analysis of Forage Sample Spectra: This graph illustrates the mean spectral signatures of forage samples (n = 600) as measured by three different scanners-TrinamiX (red line - static scan), AgroCares (green line - static scan; blue line - moving scan), and NEOSpectra (yellow - static scan; cyan - moving scan; magenta - turntable scan)-utilizing varying methods. Each line represents the average log(1/R) value across a range of wavelengths from 1400 to 2600 nm. The hue of each line represents the range between the maximum and minimum measured for each instrument.

2.3.2 Calibration Results

The calibration outcomes are summarized in Table 2.5. The NEOSpectra device in turntable mode yielded superior fit models for the calibration set variables ADF, ADL, CP, and aNDF. With respect to IVTD and NDFD, this device also demonstrated superior performance in certain metrics while remaining competitive in others. An observation is that both the moving and turntable scanning modes achieved the best calibration results when utilizing a greater number of latent variables, suggesting that dynamic scanning captures more relevant data for model calibration.

Figure 2.2 illustrates the relationship between the number of latent variables (LVs) and the root mean square error (RMSE) for various instruments and target variables. The analysis reveals that dynamic scans, represented by moving and turntable configurations, exhibit less sensitivity to larger number of latent variables. In contrast, static scans demonstrate signs of overfitting when more than 10 latent variables are used, as indicated by the deterioration in CV performance, evidenced by increasing the RMSE, despite improvements in calibration RMSE. This observation underscores the importance of scan dynamics in mitigating overfitting and enhancing the predictive accuracy of models across different latent variable configurations. The Explained Variability of the LVs can be found in Figure A.1 in the appendix. Overall, the improvement using more latent variables are not that significant for more than 10 LVs, achieving less than one percent improvement in RMSE.

2.3.3 Validation Results

The validation results are reported in Table 2.6 and summarized in Table 2.7, the performance of the NEOSpectra and Trinamix instruments across various modes and variables is quantified through metrics such as RMSE, SE, R^2, and RPD. For the NEOSpectra instrument, when operating in 'Moving' mode, CP predictions were moderately successful ($R^2 = 0.892$) with a corresponding RPD of 3.042, leading to a 'Good' classification. However, the same instrument's performance predicting IVTD in the same mode was not

Table 2.5: Statistical performance metrics for calibration models using 540 samples undried and unground alfalfa samples across different instruments and modes. The table lists the RMSE, Bias, SE, SE_{CV}, R^2, R^2_{CV}, RPD, RPD_{CV}, and the number of Latent Variables (LVs) for the variables ADF, ADL, CP, IVTD, and NDFD.

Instrument	Mode	Variable	RMSE	Bias	SE	SE_{CV}	R^2	$\text{R}^2{}_{CV}$	RPD	RPD_{CV}	LVs
AgroCares	Static	ADF	2.754	0.000	2.756	3.379	0.771	0.655	2.090	1.703	10
	Moving	ADF	1.959	0.000	1.961	2.608	0.884	0.795	2.937	2.208	15
NEOSpectra	Static	ADF	2.463	0.000	2.465	2.885	0.817	0.749	2.336	1.996	12
	Moving	ADF	2.122	0.000	2.124	2.544	0.864	0.805	2.711	2.264	20
	Turntable	ADF	**1.861**	0.000	**1.862**	**2.198**	**0.895**	**0.854**	**3.093**	**2.620**	19
Trinamix	Static	ADF	2.261	0.000	2.263	2.662	0.846	0.786	2.545	2.163	13
AgroCares	Static	ADL	1.365	0.000	1.367	1.674	0.591	0.386	1.564	1.277	11
	Moving	ADL	1.293	0.000	1.294	1.559	0.633	0.468	1.651	1.371	10
NEOSpectra	Static	ADL	1.369	0.000	1.371	1.592	0.589	0.445	1.559	1.342	11
	Moving	ADL	1.242	0.000	1.244	1.460	0.661	0.533	1.718	1.464	18
	Turntable	ADL	**1.175**	0.000	**1.176**	**1.450**	**0.697**	**0.539**	**1.817**	**1.473**	20
Trinamix	Static	ADL	1.405	0.000	1.406	1.608	0.567	0.434	1.520	1.329	10
AgroCares	Static	CP	2.010	0.000	2.012	2.306	0.792	0.727	2.195	1.915	9
	Moving	CP	1.670	0.000	1.672	1.872	0.857	0.820	2.641	2.358	8
NEOSpectra	Static	CP	1.843	0.000	1.845	2.140	0.825	0.765	2.393	2.063	11
	Moving	CP	1.513	0.000	1.514	1.799	0.882	0.834	2.916	2.454	20
	Turntable	CP	**1.328**	0.000	**1.329**	**1.601**	**0.909**	**0.869**	**3.322**	**2.758**	20
Trinamix	Static	CP	1.643	0.000	1.645	1.900	0.861	0.815	2.684	2.324	12
AgroCares	Static	IVTD	4.279	0.000	4.283	5.016	0.660	0.533	1.714	1.463	9
	Moving	IVTD	3.465	0.000	3.469	**4.177**	0.777	**0.677**	2.117	**1.758**	10
NEOSpectra	Static	IVTD	4.114	0.000	4.118	4.632	0.686	0.602	1.783	1.585	10
	Moving	IVTD	3.665	0.000	3.668	4.455	0.750	0.632	2.002	1.648	19
	Turntable	IVTD	**3.409**	0.000	**3.412**	4.195	**0.784**	0.674	**2.152**	1.750	20
Trinamix	Static	IVTD	3.757	0.000	3.760	4.372	0.738	0.645	1.953	1.679	12
AgroCares	Static	NDFD	6.534	0.000	6.540	7.744	0.524	0.333	1.450	1.224	10
	Moving	NDFD	5.653	0.000	5.658	6.862	0.644	0.476	1.676	1.382	10
NEOSpectra	Static	NDFD	6.317	0.000	6.323	7.173	0.555	0.428	1.500	1.322	11
	Moving	NDFD	5.730	0.000	5.735	6.980	0.634	0.458	1.653	1.358	19
	Turntable	NDFD	5.485	0.000	5.490	**6.508**	0.665	**0.529**	1.727	**1.457**	19
Trinamix	Static	NDFD	**5.390**	0.000	**5.395**	7.081	**0.676**	0.442	**1.757**	1.339	20
AgroCares	Static	aNDF	3.811	0.000	3.814	4.370	0.863	0.820	2.700	2.356	10
	Moving	aNDF	3.256	0.000	3.259	3.829	0.900	0.862	3.159	2.689	8
NEOSpectra	Static	aNDF	3.243	0.000	3.246	3.752	0.901	0.867	3.172	2.744	15
	Moving	aNDF	3.007	0.000	3.010	3.573	0.915	0.880	3.421	2.881	20
	Turntable	aNDF	**2.605**	0.000	**2.608**	**3.031**	**0.936**	**0.913**	**3.949**	**3.397**	20
Trinamix	Static	aNDF	2.905	0.000	2.908	3.828	0.920	0.862	3.541	2.689	20

SE - Standard Error, RPD - Ratio of Performance to Deviation, CV - Cross-Validated, LVs - Latent Variables, ADF - Acid Detergent Fiber, ADL - Acid Detergent Lignin, CP - Crude Protein, IVTD - in-vitro Total Digestibility, NDFD - Neutral Detergent Fiber Digestibility, aNDF - actual Neutral Detergent Fiber

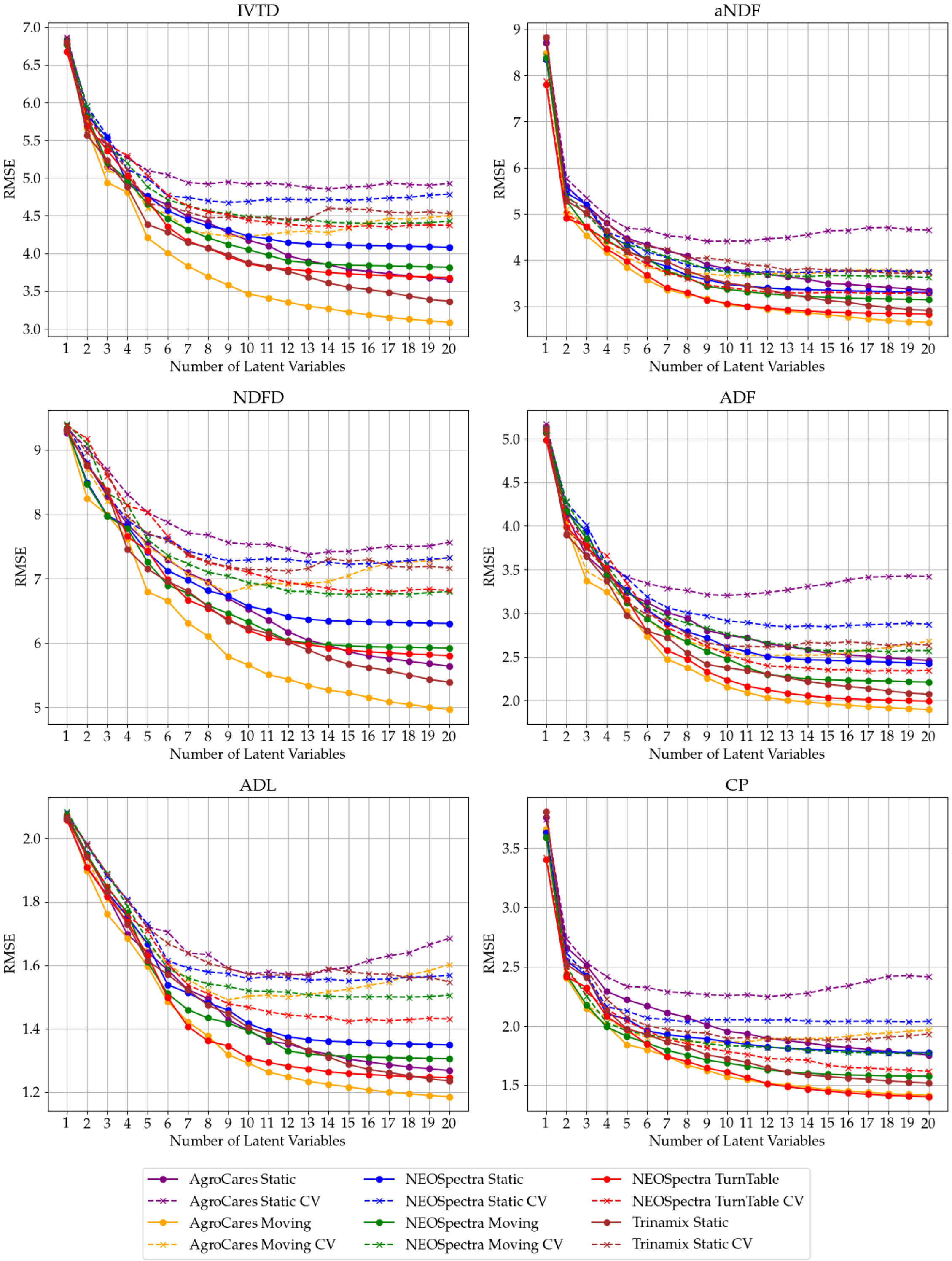


Figure 2.2: RMSE vs Latent Variables for each Variable. This figure shows the root mean squared error (RMSE) values for different numbers of latent variables across various instruments and target variables. The RMSE values for both calibration and cross-validation (CV) are plotted for each instrument, differentiated by color (purple - AgroCares Static, orange - Agrocares Moving, blue - NEOSpectra Static, green - NEOSpectra Moving, red - NEOSpectra Turntable, brown - Trinamix Static) and line style (continuous - callibration, dashed - CV).

useful, with a lower R^2 of 0.743 and an RPD of 1.974, reflecting a 'Very poor' classification. When utilizing the 'Turntable' mode, predictions of ADF and ADL yielded a 'Fair' and 'Very poor' classification, respectively, indicating a varied efficacy based on the forage constituent analyzed. In contrast, the Trinamix instrument in 'Static' mode demonstrated 'Successful' prediction for aNDF with an R^2 of 0.916 and an RPD of 3.452, garnering a 'Very good' classification. These results reflect the nuanced capabilities of each instrument and mode combination, emphasizing the importance of selecting the appropriate setup for specific analytical needs in forage assessment. Figure 2.3 makes it clear that the calibrated models did not perform well in predicting NDFD, ADL, and IVTD on the validation set for undried haylage samples.

When calibrated exclusively with static scans, TrinamiX had a better RMSE, SE and RPD than the NEOSpectra for predicting CP and ADL. These findings concur with the insights of Feng et al. [11], highlighting the enhanced spectral quality afforded by moving scans due to their capacity to encapsulate a more generalized and homogeneous representation of the samples. A closer examination of the calibration data delineated in table 2.5 reveals a performance hierarchy within the same instrument, with the order of efficacy being Turntable > Moving > Static. This sequence also correlates with the increasing number of latent variables that can be utilized in the NEOSpectra scanning process, thereby suggesting that sliding scans not only improve the spectral quality but also allow for a better model calibration.

Utilizing the same scanning pattern—whether static, moving, or turntable—tends to yield comparable calibration performance across different devices, as evidenced by the minimal variation in the R^2 and the RPD values, usually within the same class range of success, according to table 2.2. This is consistent across most variables, with the notable exception of Neutral Detergent Fiber (aNDF), where the Trinamix instrument in static mode achieved an RPD of 3.453, surpassing those of AgroCares at 2.506 and NEOSpectra at 2.864. The findings thus suggest that the methodology of spectral data acquisition is more important than the choice of handheld instrument.

Table 2.6: Validation performance metrics for different instruments operating in Static, Moving, and Turntable modes. The metrics include RMSE, Bias, SE, R^2, Slope, Intercept, and RPD for the validation of variables ADF, ADL, CP, IVTD, NDFD, and aNDF using a set of 60 samples. This table facilitates the comparison of model precision and prediction accuracy across diverse instruments and scanning configurations for the validation dataset.

Instrument	Mode	Variable	RMSE	Bias	SE	R^2	Slope	Intercept	RPD
AgroCares	Static	ADF	2.949	-0.150	2.970	0.761	0.973	1.191	2.047
	Moving	ADF	3.015	-0.654	2.968	0.751	0.928	3.386	2.003
NEOSpectra	Static	ADF	2.490	-0.203	2.502	0.830	1.026	-0.805	2.425
	Moving	ADF	2.283	-0.446	2.258	0.857	0.923	3.374	2.645
	Turntable	ADF	**2.207**	-0.193	**2.217**	**0.866**	0.946	2.269	**2.736**
Trinamix	Static	ADF	2.536	-0.332	2.536	0.824	1.090	-3.124	2.381
AgroCares	Static	ADL	2.400	-0.040	2.420	0.109	0.644	2.705	1.059
	Moving	ADL	2.247	-0.251	2.252	0.219	0.835	1.450	1.132
NEOSpectra	Static	ADL	2.050	-0.079	2.066	0.350	0.963	0.356	1.240
	Moving	ADL	2.013	-0.321	2.004	0.373	1.018	0.194	1.263
	Turntable	ADL	**1.794**	-0.178	**1.800**	**0.502**	1.040	-0.115	**1.417**
Trinamix	Static	ADL	1.961	-0.307	1.953	0.405	1.257	-1.550	1.297
AgroCares	Static	CP	2.003	0.356	1.988	0.783	0.893	1.408	2.144
	Moving	CP	1.729	0.329	1.712	0.838	0.937	0.700	2.484
NEOSpectra	Static	CP	1.977	0.333	1.965	0.788	0.997	-0.288	2.172
	Moving	CP	**1.412**	0.074	**1.422**	**0.892**	0.986	0.150	**3.042**
	Turntable	CP	1.517	0.140	1.524	0.875	0.935	0.911	2.831
Trinamix	Static	CP	1.712	0.101	1.723	0.841	0.925	1.112	2.509
AgroCares	Static	IVTD	4.558	-0.709	4.540	0.577	0.862	11.412	1.537
	Moving	IVTD	4.141	0.010	4.176	0.651	0.857	11.125	1.692
NEOSpectra	Static	IVTD	4.040	-0.315	4.061	0.668	0.911	7.270	1.735
	Moving	IVTD	**3.549**	-0.262	3.570	**0.743**	0.927	5.942	**1.974**
	Turntable	IVTD	3.550	-0.277	**3.569**	0.743	0.854	11.650	1.974
Trinamix	Static	IVTD	4.142	-0.180	4.173	0.651	0.912	7.068	1.692
AgroCares	Static	NDFD	8.355	-1.046	8.360	-0.019	0.497	30.314	0.991
	Moving	NDFD	7.406	-0.232	7.465	0.200	0.645	21.191	1.118
NEOSpectra	Static	NDFD	6.515	-0.590	6.543	0.381	0.838	10.082	1.271
	Moving	NDFD	5.943	-0.164	5.991	0.485	0.952	3.016	1.393
	Turntable	NDFD	**5.544**	-0.348	**5.579**	**0.552**	0.816	11.202	**1.493**
Trinamix	Static	NDFD	7.295	-0.515	7.338	0.223	0.649	21.140	1.135
AgroCares	Static	aNDF	4.380	-0.501	4.388	0.841	1.018	-0.462	2.506
	Moving	aNDF	3.808	-0.865	3.739	0.880	1.020	-0.186	2.883
NEOSpectra	Static	aNDF	3.832	-0.713	3.797	0.878	1.050	-1.935	2.864
	Moving	aNDF	3.494	-0.174	3.519	0.899	0.991	0.666	3.141
	Turntable	aNDF	3.304	-0.404	3.307	0.909	0.991	0.872	3.322
Trinamix	Static	aNDF	**3.180**	0.189	**3.201**	**0.916**	1.010	-0.705	**3.452**

SE - Standard Error, RPD - Ratio of Performance to Deviation, ADF - Acid Detergent Fiber, ADL - Acid Detergent Lignin, CP - Crude Protein, IVTD - in-vitro Total Digestibility, NDFD - Neutral Detergent Fiber Digestibility, aNDF - actual Neutral Detergent Fiber

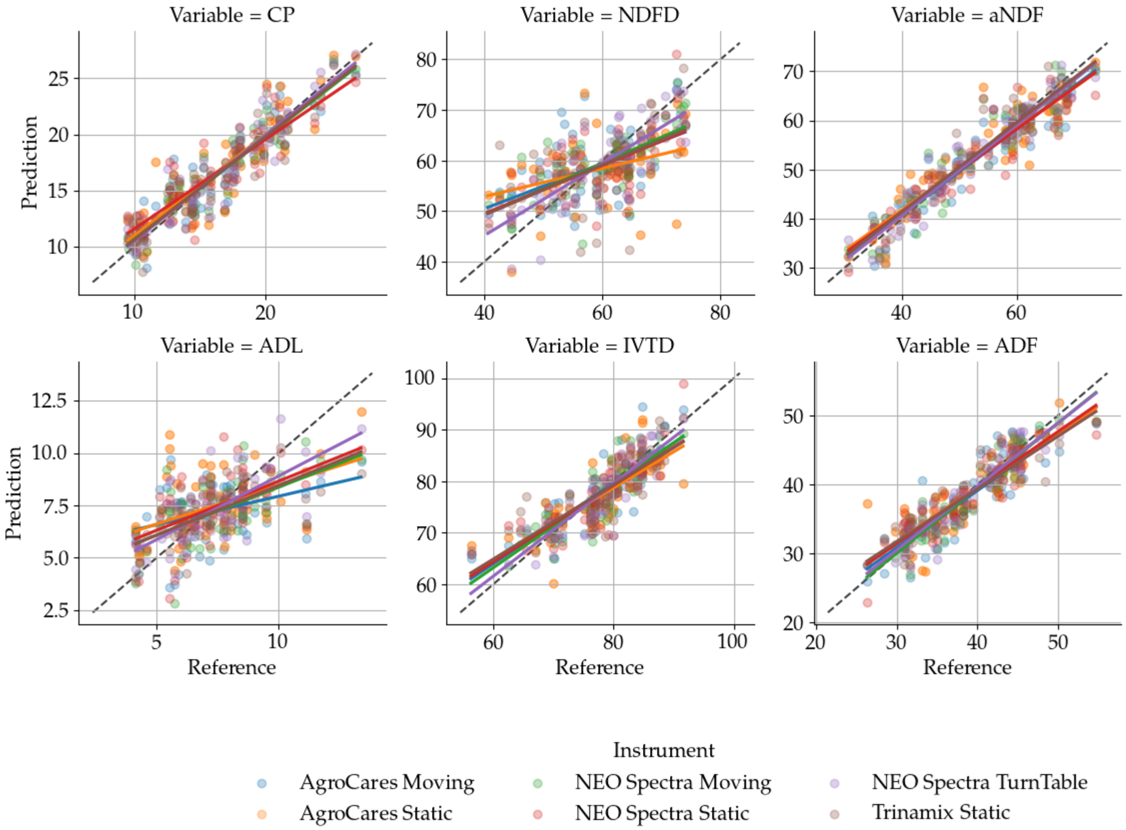


Figure 2.3: Comparative evaluation of three handheld spectrometers and methods used to predict nutritional content in feed samples. The different colors and shapes represent readings from moving, static, or turntable methods of using the AgroCares, NEO Spectra, and TrinamiX instruments. Each dot represents the pair of reference data and the prediction using the calibrated PLSR model from the validation set (n = 60). The regression lines for each method showcase the accuracy and precision in predicting the content of Crude Protein (CP), Neutral Detergent Fiber (aNDF), Acid Detergent Fiber (ADF), Acid Detergent Lignin (ADL), Neutral Detergent Fiber Digestibility (NDFD), and In Vitro True Digestibility (IVTD). The dashed black line represents a 1:1 agreement between the reference and predicted values.

Table 2.7: Performance of the best model for each predicted variable on the validation set.

Instrument	Mode	Variable	RMSE	SE	R^2	RPD	Success (R^2[14])	Classification (RPD [24])
NEOSpectra	Moving	CP	1.412	1.422	0.892	3.042	Moderately Successful	Good
	Moving	IVTD	3.549	3.570	0.743	1.974	Not Useful	Very poor
	Turntable	ADF	2.207	2.217	0.866	2.736	Moderately Successful	Fair
	Turntable	ADL	1.794	1.800	0.502	1.417	Not Useful	Very poor
	Turntable	NDFD	5.544	5.579	0.552	1.493	Not Useful	Very poor
Trinamix	Static	aNDF	3.180	3.201	0.916	3.452	Successful	Very good

RMSE - Root Mean Squared Error, SE - Standard Error, RPD - Ratio of Performance to Deviation, ADF - Acid Detergent Fiber, ADL - Acid Detergent Lignin, CP - Crude Protein, IVTD - in-vitro Total Digestibility, NDFD - Neutral Detergent Fiber Digestibility, aNDF - actual Neutral Detergent Fiber

Figure 2.4 presents normalized boxplots of prediction errors, facilitating a more nuanced comparison of calibration performance. Notably, CP, ADF, and aNDF demonstrate the most favorable results, characterized by minimal bias and RMSE, with most prediction errors falling within one standard deviation. Conversely, ADL predictions were less accurate, exhibiting a multitude of outliers, as reflected in Table 2.7, indicating a disparity between the predicted and observed values.

The standard error of laboratory (SEL) values reported by Cherney et al. (2021) [8] for the wet chemistry of the same samples indicated errors that are an order of magnitude lower than the root mean square error (RMSE) values from our results presented in Table 2.7 ($SEL_{aNDF} \leq$ 0.66, $SEL_{ADF} \leq$ 0.70, $SEL_{ADL} \leq$ 0.30, $SEL_{IVTD} \leq$ 0.76, $SEL_{NDFD} \leq$ 2.36, and $SEL_{CP} \leq$ 0.44).

Compared to existing literature, our calibration results are compatible with previously reported findings. For undried samples, [8] evaluated the Aurora instrument for haylage, corn silage, and Total Mixed Ration, while[11] assessed the NEOSpectra for corn silage, alfalfa, grass, and mixed alfalfa and grass silage. Both studies utilized moving scans. As shown in Figure 2.5, our results with haylage samples exhibit similar characteristics to the best metrics from the literature, further supporting the notion that dynamic scans often outperform other methods. Additionally, our findings indicate that ADF, aNDF, and CP achieve moderately successful to excellent calibrations, whereas IVTD, NDFD, and ADL

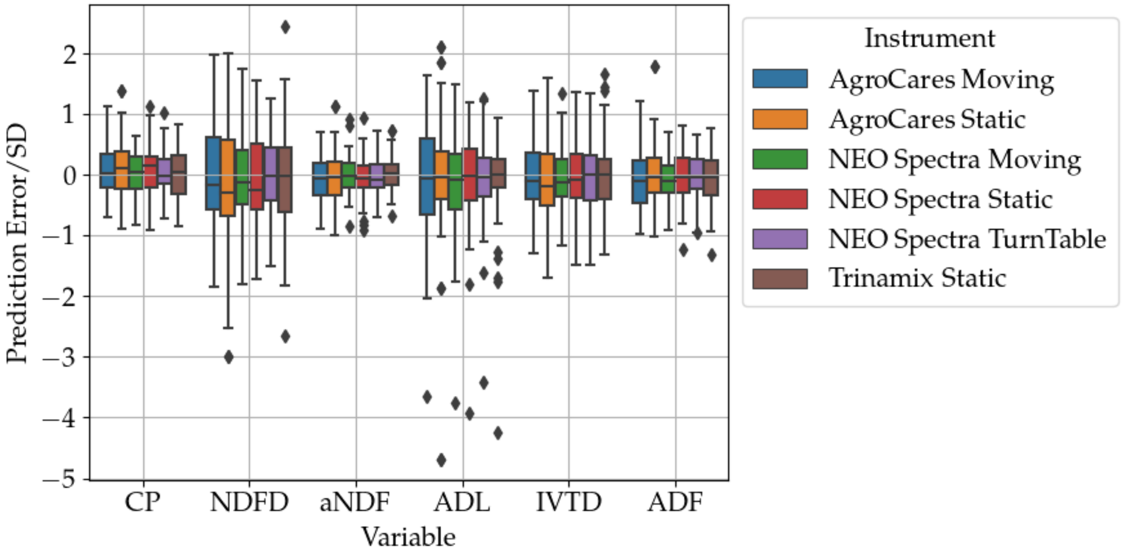


Figure 2.4: The chart presents the normalized distribution of prediction errors on the validation set for six forage quality variables—CP, NDFD, aNDF, ADL, IVTD, and ADF—obtained using different spectral scanning instruments and methods. Each boxplot shows the median, quartiles, and outliers for the prediction error standard deviation (SD) of each method.

do not yield useful models according to [14] R^2 criteria.

As one of the goals of portable NIRS is to be used for in-field forage quality estimation, it is important to understand how the prediction model performance is affected when using undried samples. Since water has high absorption bands at 1400 and 1900 nm, it creates interference in the raw spectrum of the material in the NIR region [23]. Figure 2.6 illustrates how our model performance compares to models calibrated on dried samples. It is evident that the water content of the samples affects the performance of the models to varying degrees for all forage quality metrics studied. ADF, aNDF, and CP are less impacted by water, exhibiting similar R^2 values to those of dried materials. In contrast, IVTD, NDFD, and ADL are severely affected by the water content, resulting in lower R^2 values. Further investigation into how water influences the PLSR loading factors is necessary to better understand whether these effects can be mitigated.

Figures 2.7 and 2.8 display the first two latent variables that contain most of the explained variance of the PLS models (Appendix A.1). It is evident that the main absorption

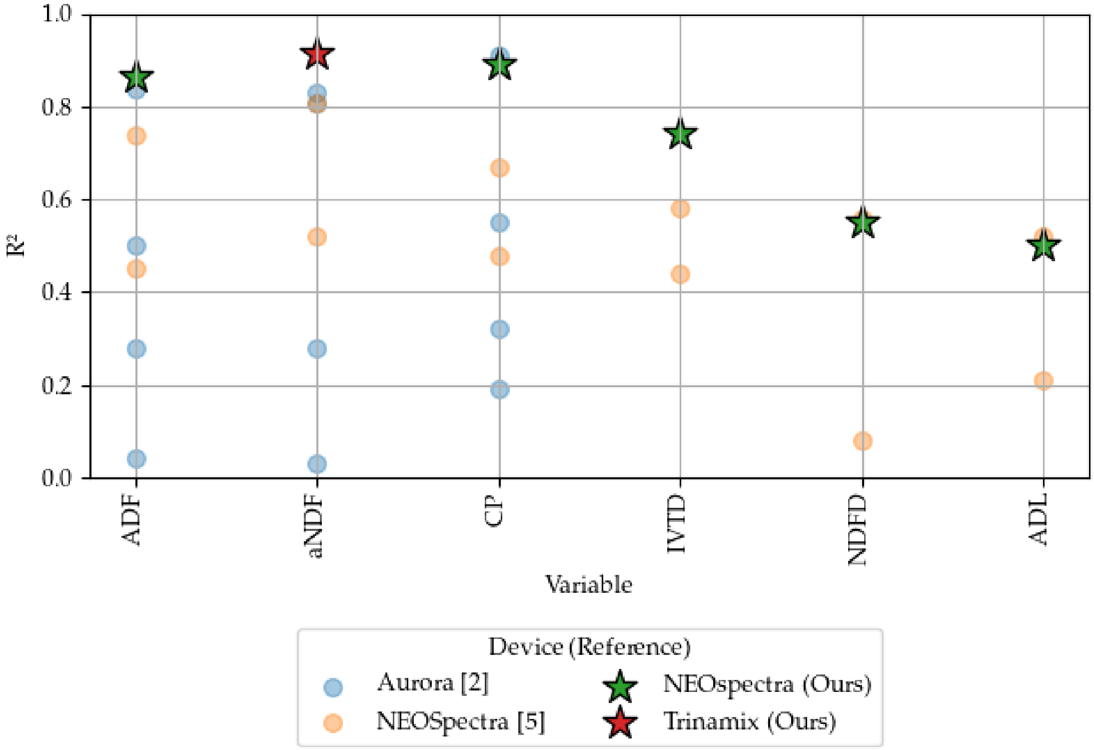


Figure 2.5: Undried data R^2 comparison with Aurora [8], calibrated for haylage, corn silage, and total mixed ration,and NEOSpectra [11] calibrated for grass, alfalfa, and mixed silage forages. Both references were sampled using moving scans. The dots represents the metrics obtained by the references and the stars represents the metrics obtained by our best model.

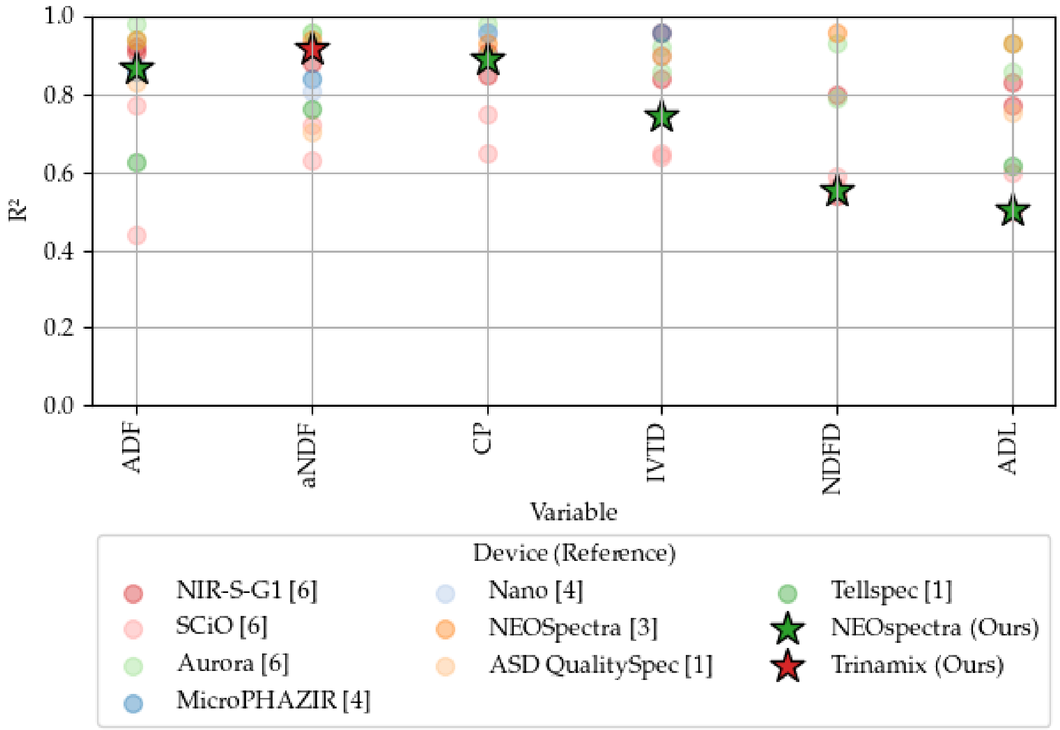


Figure 2.6: Comparison of R^2 values from models calibrated on dried samples (literature) versus our model calibrated on undried samples. Tellspec and ASD QualitySpec [19] were calibrated for grass. NEOSpectra [10] was calibrated for grass, alfalfa, and mixed silage forages. Nano and MicroPHAZIR [1] were calibrated for grass forages. NIR-S-G1, SCiO, and Aurora [6] were calibrated for alfalfa and grass forages. The dots represent the metrics obtained from the references, and the stars represent the metrics obtained by our best model.

bands of water play a role in the loading factors of the latent variables, particularly at the 1904 nm band, where water has a significant absorption peak. These effects arise from the interactions of water with the O-H groups present in carbohydrates, fats, and proteins, which can form hydrogen bonds with most types of fiber. These results are consistent with findings obtained from studies on small grains [23].

Collectively, these analyses underscore the significance of the scanning pattern over the specific technology or instrument used. The consistency in data acquisition methodology emerges as a critical factor in the calibration performance, influencing the robustness of predictive models more substantially than the hardware utilized. Furthermore, the lower performance of certain variables is likely due to the water's electromagnetic absorption and interaction with the undried sample material.

2.4 Conclusions

This study systematically explores the influence of scanning methodology and instrument design on the efficacy of spectroscopic models in forage analysis. Our findings are derived from a set of 600 ensiled forage samples collected across New York state. As detailed in tables 2.5 and 2.6, and summarized in Table 2.7, highlight the importance of the spectral acquisition technique plays over the specific technical features of handheld NIRS devices.

The consistency observed across the instruments when identical scanning patterns were employed underscores the methodological influence over technology. Specifically, the NEOSpectra instrument, when employed in a dynamic mode, demonstrates a significant advantage in the predictive accuracy for all variables. This suggests that the precision and reliability of predictions are more heavily contingent upon applying robust and consistent scanning protocols.

Based on the analysis of Figure 2.2 and Table A.1, we recommend limiting the number of latent variables to 7-10 to avoid overfitting and ensure future model performance. Although our results points that 11-20 LVs can have a small improvement in the unseen validation set, the explained variance of these variables does not improve significantly the results to

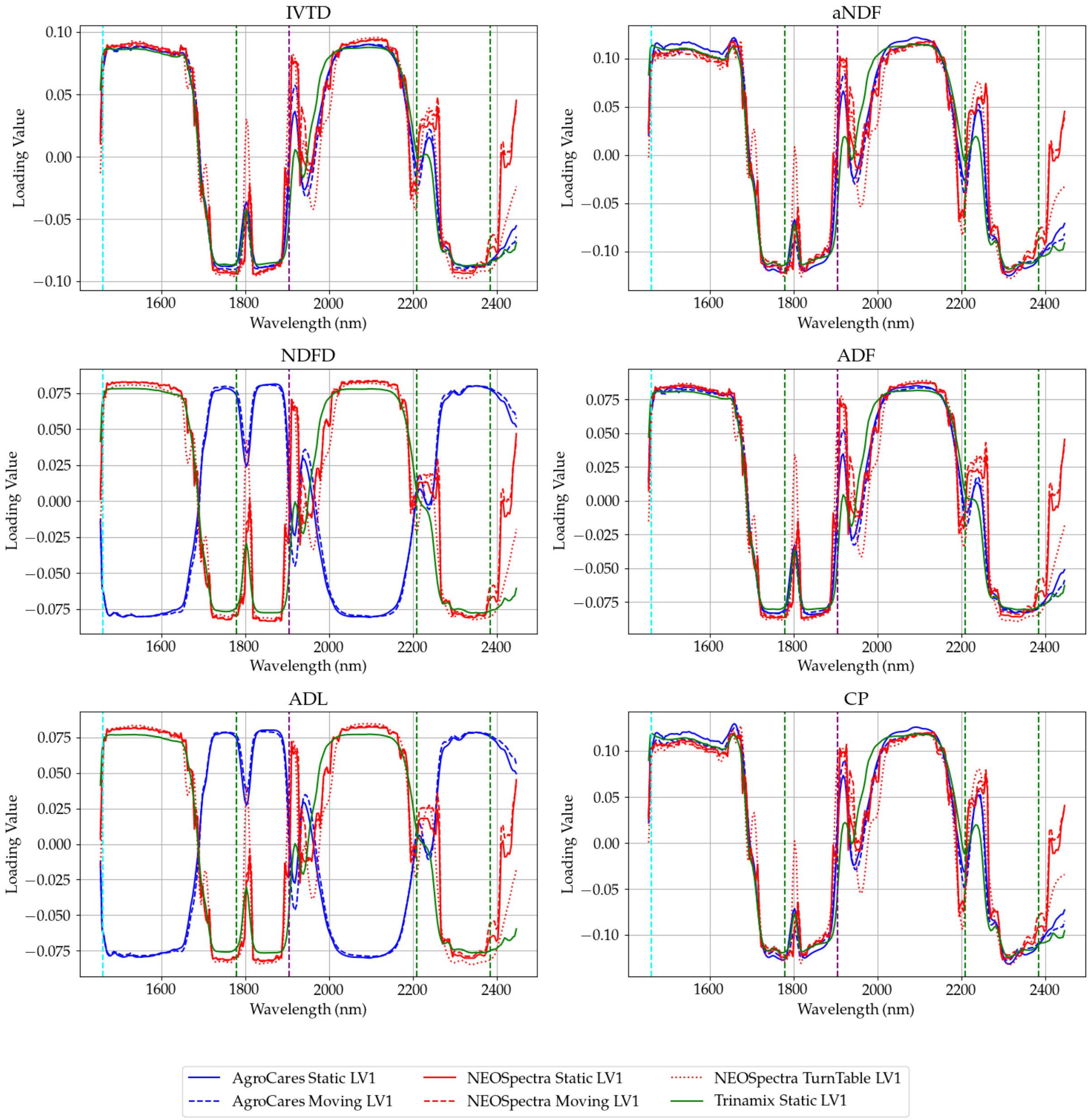


Figure 2.7: Loading values of the first latent variable of the spectrum (first derivative). Instruments are divided by color (AgroCares - blue, NEOSpectra - red, Trinamix - green). The scan mode are divided by the line style (continuous - static, dashed - moving, dotted - turntable). The vertical lines are the water absorption bands. Very small absorption bands (1778, 2208, and 2384 nm) in green. Large absorption band (1460 nm) is illustrated in cyan. Very large absorption band (1904 nm) in purple.

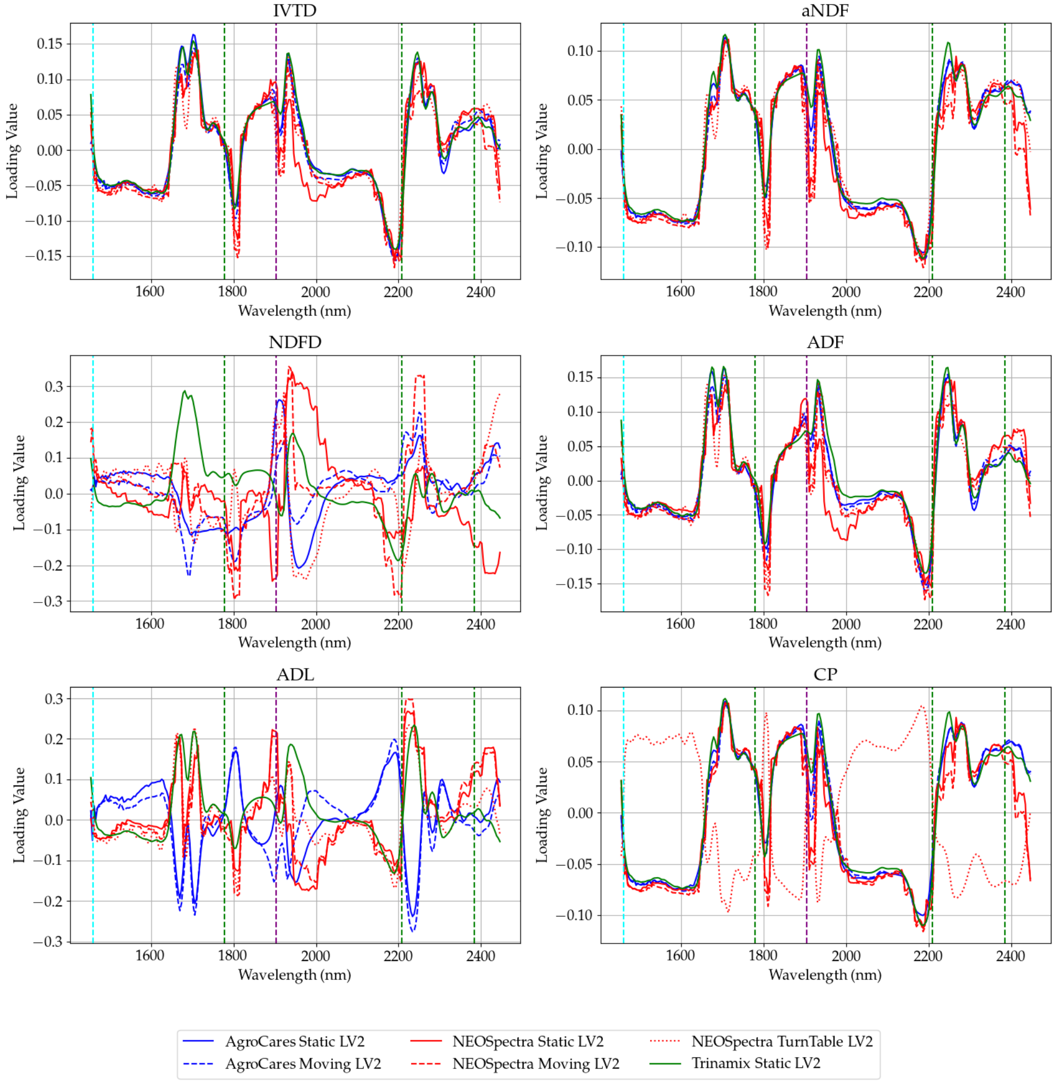


Figure 2.8: Loading values of the second latent variable of the spectrum (first derivative). Instruments are divided by color (AgroCares - blue, NEOSpectra - red, Trinamix - green). The scan mode are divided by the line style (continuous - static, dashed - moving, dotted - turntable). The vertical lines are the water absorption bands. Very small absorption bands (1778, 2208, and 2384 nm) in green. Large absorption band (1460 nm) is illustrated in cyan. Very large absorption band (1904 nm) in purple.

justify using them. Therefore, a careful balance must be struck between model complexity and predictive stability.

Comparative analyses, particularly for undried forage, have aligned with findings from previous literature, confirming the validity of our models within the expected performance parameters. Moreover, the results have revealed that scanning modes incorporating movement tend to enhance the homogeneity of the sample representation, which is critical in achieving high-quality spectral data. When comparing dried and grounded material calibrations, we can see that fiber constituents and protein are less impacted by water absorption. However, there remains a knowledge gap in understanding the water interactions of the undried and unground forage constituents, specifically how water interaction affects their NIR spectral characteristics. The loadings obtained through partial least squares regression of the NIR spectra highlight the critical role of variance at wavelengths associated with O-H absorptions in constructing models for these materials. The behavior of water within complex agricultural substances is expected to differ from that of liquid water.

Figure 2.3 and Figure 2.4 provide visual confirmation of the comparative and error distribution analyses, respectively, illustrating the nuanced performance across different forage constituents and underscoring the models that exhibit both high accuracy and those with room for improvement

In summary, this research affirms the importance of the scanning pattern in developing robust near-infrared spectroscopic models. It contributes valuable insights that may guide practitioners in selecting the most suitable instruments and modes for forage quality assessment. As the field advances, future studies should further refine these methodologies, optimizing the balance between technological innovation and practical application for enhanced forage analysis. Evaluation of embedded NIR sensors in agricultural machinery to predict forage quality and properties is one of the path forward to undried forage research.

References

[1] J. J. Acosta, M. S. Castillo, and G. R. Hodge. “Comparison of benchtop and handheld near-infrared spectroscopy devices to determine forage nutritive value”. en. In: *Crop Sci.* 60.6 (Nov. 2020), pp. 3410–3422.

[2] F. Barton and D. Burdick. “Prediction of forage quality with NIR reflectance spectroscopy”. In: *Proceedings Of The Xiv International Grassland Congress.* CRC Press. 2019, pp. 532–534.

[3] K. B. Beć, J. Grabska, and C. W. Huck. “Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers”. en. In: *Chemistry* 27.5 (Jan. 2021), pp. 1514–1532.

[4] K. B. Beć et al. “Handheld near-infrared spectrometers: Where are we heading?” In: *NIR news* 31.3-4 (June 2020), pp. 28–35.

[5] L. Bedord. *John Deere Adds Manure Constituent Sensing to HarvestLab 3000.* `https://www.agriculture.com/news/technology/john-deere-adds-manure-constituent-sensing-to-harvestlab-3000`. Accessed: 2024-3-6. June 2019.

[6] P. Berzaghi, J. H. Cherney, and Casler. “Prediction performance of portable near infrared reflectance instruments using preprocessed dried, ground forage samples”. In: *Comput. Electron. Agric.* 182 (Mar. 2021), p. 106013.

[7] L. Chen, L. Xing, and L. Han. “Review of the application of near-infrared spectroscopy technology to determine the chemical composition of animal manure”. en. In: *J. Environ. Qual.* 42.4 (July 2013), pp. 1015–1028.

[8] J. H. Cherney, M. F. Digman, and D. J. Cherney. “Handheld NIRS for forage evaluation”. In: *Comput. Electron. Agric.* 190 (Nov. 2021), p. 106469.

[9] D. Cozzolino and M. Labandera. “Determination of dry matter and crude protein contents of undried forages by near-infrared reflectance spectroscopy”. en. In: *J. Sci. Food Agric.* 82.4 (Mar. 2002), pp. 380–384.

[10] M. F. Digman, J. H. Cherney, and D. J. R. Cherney. “The Relative Performance of a Benchtop Scanning Monochromator and Handheld Fourier Transform Near-Infrared Reflectance Spectrometer in Predicting Forage Nutritive Value”. en. In: *Sensors* 22.2 (Jan. 2022).

[11] X. Feng et al. “Practical Considerations for Using the NeoSpectra-Scanner Handheld Near-Infrared Reflectance Spectrometer to Predict the Nutritive Value of Undried Ensiled Forage”. en. In: *Sensors* 23.4 (Feb. 2023).

[12] G. Gislon et al. “Forage systems and sustainability of milk production: Feed efficiency, environmental impacts and soil carbon stocks”. en. In: *J. Clean. Prod.* 260.121012 (July 2020), p. 121012.

[13] L. Hart et al. “Comparison of Spectral Reflectance-Based Smart Farming Tools and a Conventional Approach to Determine Herbage Mass and Grass Quality on Farm”. en. In: *Remote Sensing* 12.19 (Oct. 2020), p. 3256.

[14] D. F. Malley et al. “Compositional Analysis of Cattle Manure During Composting Using a Field-Portable Near-Infrared Spectrometer”. In: *Commun. Soil Sci. Plant Anal.* 36.4-6 (Mar. 2005), pp. 455–475.

[15] D. Mendes de Oliveira et al. “Comparative analysis of compact and benchtop near-infrared spectrometers for forage nutritional trait measurements”. en. In: *Microchem. J.* 196.109682 (Nov. 2023), p. 109682.

[16] J. Muncan and R. Tsenkova. “Aquaphotomics-from innovative knowledge to integrative platform in science and technology”. en. In: *Molecules* 24.15 (July 2019), p. 2742.

[17] G. Rego et al. “A portable IoT NIR spectroscopic system to analyze the quality of dairy farm forage”. In: *Comput. Electron. Agric.* 175 (Aug. 2020), p. 105578.

[18] R. Roelofs et al. “A meta-analysis of overfitting in machine learning”. In: *Adv. Neural Inf. Process. Syst.* (2019), pp. 9175–9185.

[19] I. R. Rukundo et al. “Comparing the use of handheld and benchtop NIR spectrometers in predicting nutritional value of forage”. en. In: *Appl. Eng. Agric.* 37.1 (2021), pp. 171–181.

[20] J. S. Shenk, M. O. Westerhaus, and M. R. Hoover. “Analysis of Forages by Infrared Reflectance1”. In: *J. Dairy Sci.* 62.5 (May 1979), pp. 807–812.

[21] R. Tsenkova et al. “Essentials of aquaphotomics and its chemometrics approaches”. en. In: *Front. Chem.* 6 (Aug. 2018), p. 363.

[22] M. E. Valentine et al. “Comparison of in vitro long digestion methods and digestion rates for diverse forages”. en. In: *Crop Sci.* 59.1 (Jan. 2019), pp. 422–435.

[23] P. Williams. “Influence of water on prediction of composition and quality factors: The aquaphotomics of low moisture agricultural materials”. en. In: *J. Near Infrared Spectrosc.* 17.6 (Dec. 2009), pp. 315–328.

[24] P. Williams. “The RPD Statistic: A Tutorial Note”. In: *NIR news* 25.1 (Feb. 2014), pp. 22–26.

[25] H. Yan et al. “Handheld Near-Infrared Spectroscopy: State-of-the-Art Instrumentation and Applications in Material Identification, Food Authentication, and Environmental Investigations”. en. In: *Chemosensors* 11.5 (May 2023), p. 272.

Chapter 3

Pre-harvest loss quantification in grain crops

William Yamada, Jordi Francis, Jack Jewison, Troy Runge and Matthew Digman

Publication note: This chapter is based on the following article published in *Computers and Electronics in Agriculture*:

Yamada, W., Francis, J., Jewison, J., Runge, T., & Digman, M. (2025). *Pre-harvest loss quantification in grain crops.* **Computers and Electronics in Agriculture**, 236, 110405. `https://doi.org/10.1016/j.compag.2024.110405`

The author of this thesis led the project design, image analysis pipeline, model training, and writing of the manuscript.

Abstract: Traditional methods for measuring pre-harvest loss, such as using quadrats, are labor-intensive and provide sparse data coverage.This study proposes an automated approach that leverages computer vision to replace and enhance the current method, using advanced imaging technologies and deep learning methodologies to detect and quantify pre-harvest losses in grain crops. Specifically, the methodology employs a camera mounted on the front snout of a ground vehicle, allowing continuous image capture along the crop rows. By automating image collection and analysis, this approach provides denser spatial coverage across the field, reduces errors associated

with manual sampling and human judgment, and significantly accelerates the process compared to traditional quadrat sampling. In addition, this approach offers the potential to segregate different types of pre-harvest loss, such as natural shattering versus losses caused by mechanical disturbance, providing a level of granularity not achievable with conventional quadrat methods. By leveraging state-of-the-art object detection architectures the system is designed to handle the complex visual environment of the field floor, where grains may be obscured by crop residue, shadows, and similarly-colored objects such as stones. This capability represents a significant advancement over traditional methods, which cannot distinguish between these different loss sources. The images were annotated using the Segment Anything Model (SAM) to ensure consistency and accuracy across the dataset.

Several state-of-the-art models were trained and evaluated on the collected data, including Mask RCNN, YOLOX, DETR, and a modified YOLOv8-p2. The modified YOLOv8-p2 model, which incorporated a p2 head to improve the detection of smaller objects, outperformed the others, yielding the highest Precision, Recall, and F1 scores on both the soybean (Precision=0.727, Recall=0.694, F1=0.710) and wheat (Precision=0.709, Recall=0.688, F1=0.698) datasets. Integrating the 850 nm NIR image channel did not produce a meaningful boost in performance, as evidenced by the soybean (Precision=0.741, Recall=0.689, F1=0.715) and wheat (Precision=0.729, Recall=0.690, F1=0.709) results.

This research demonstrates that it is possible to integrate a vision system on the header of a combine and identify the initial shedding loss in the field. Future work will focus on refining the models further, exploring their applicability to other crop types, and integrating real-time processing and automation in data collection and analysis.

3.1 Introduction

Small grain crops, such as wheat, soybeans, and oats, are crucial to global food security [2]. Pre-harvest loss can result from various factors, including climatic conditions, pest infestations, diseases, and inadequate agricultural practices (e.g., improper fertilization schedules, poorly timed pesticide applications, or inadequate crop rotation practices). In

2018, an estimated 26-30% of the soybean and wheat total possible biological product was lost due to pre-harvest factors [30]. Developing better methods for quantifying these losses across the supply chain is essential for designing effective mitigation strategies, optimizing resource allocation, and reducing economic losses for farmers while improving food security.

The current standard for measuring pre-harvest loss from shedding involves using a quadrat of 1 m^2 and manually counting the loss within this area [20]. This method results in sparse measurements across the field and is labor-intensive. Additionally, in some crops, such as wheat, the physical presence of the observer in the field can cause additional losses due to plant disturbance, leading to measurement bias. In some crops, such as wheat, it is difficult to walk over the crop and measure the loss with a quadrat without causing additional losses.

Measurement of field pre-harvest shedding loss in conjunction with post-harvest loss provides an understanding of its contribution to total crop loss, which is essential to determine harvest machine loss (*viz.* header, threshing, and separating) [30]. Quantifying the loss before and after harvesting makes it possible to isolate the specific contribution of machine losses to the overall loss of the crop. This approach is crucial because post-harvest field loss has already been demonstrated to be measurable, as indicated by previous studies [32]. However, post-harvest assessments often rely on combine yield loss monitors, which estimate losses through impact sensors mounted on straw walkers and sieves. These systems provide real-time feedback to operators but are limited by sensor accuracy, field variability, and challenges in detecting unthreshed grain loss [3, 7]. Integrating pre-harvest loss measurements with post-harvest assessments allows for identifying inefficiencies and areas for improvement in harvesting technology, ultimately leading to reduced losses and increased efficiency in agricultural practices.

Despite its importance, there is a lack of standardized, efficient, and accurate methods for quantifying pre-harvest loss in small grain crops. The existing method is too labor-intensive and does not represent the variability across the field [20]. Additionally, the observer can cause measurable losses as they interact with the crop. Furthermore, current loss estimation methods, including combine yield loss monitors, do not account for pre-harvest loss, leading to potential underestimation of total yield loss. This methodology gap hinders agricultural manufacturers' ability to optimize the harvester and for farmers to understand how their management practices affect shedding losses.

Recent advancements in machine learning and computer vision in agriculture could be leveraged for predicting pre-harvest losses in small grains. Previous studies have explored the use of remote sensing and machine learning to monitor crop health [36, 44] and detect stress factors [37, 14]. Other works have demonstrated the effectiveness of deep learning models in identifying specific crop diseases [21, 25, 43] and pests[48, 12], contributing to more efficient management practices. Additionally, the integration of UAVs or ground-based systems for comprehensive field analysis has been increasingly adopted, showing promising results in precision agriculture [17, 4, 35, 38, 9]. For object detection, the fusion of RGB-NIR images has been used to increase the deep learning model's performance [10, 52, 29, 54].

This study explores the strengths and limitations of aerial and ground-based imaging for pre-harvest loss detection. Aerial imaging provides broader spatial coverage and efficient large-scale data collection, while ground-based imaging offers higher resolution at the soil surface, essential for detecting small grains obscured by crop residue or soil texture. By comparing these approaches, the study evaluates their practical feasibility, accuracy, and operational trade-offs for pre-harvest loss quantification. This directly addresses the lack of consensus in prior work [17, 4, 35, 38, 9] regarding the optimal sensing platform for field-level crop monitoring, particularly for detecting fallen grains.

Beyond these, recent advances in object detection architectures offer further potential for enhancing pre-harvest loss detection. Roy and Bhaduri [40] proposed a real-time growth stage detection model based on DenseNet-fused YOLOv4, designed to handle high degrees of crop occultation, which is particularly relevant for field-based imaging under natural occlusion. In a related work, Roy et al. [41] developed a fine-grain object detection model by modifying YOLOv4, integrating DenseNet for improved feature transfer and reuse, demonstrating superior accuracy in detecting plant diseases. Further refinement of this approach led to DenseSPH-YOLOv5, which combines DenseNet with a Swin-Transformer prediction head, incorporating attention mechanisms to enhance damage detection in complex environments [39]. Jiang et al. [23] introduced a novel multiple graph learning neural network (MGLNN), capable of learning from multiple graph structures simultaneously, highlighting the potential of graph-based approaches for handling multi-view or multi-sensor data in agricultural monitoring. These technological advancements present an opportunity to develop automated pre-harvest loss detection systems that can complement or surpass existing loss monitoring techniques.

In the context of grain detection, deep learning research has shown significant advancements in recent years. Sun et al. [46] successfully counted rice grains from panicles using YOLOv7, while Geng et al. [16] developed a U-Net-based model to count grains from wheat spikes. James et al. [22] introduced GrainPointNet, a method designed to predict grain counts from high-resolution point clouds. Lin and Guo [28] and Mosley [33] applied aerial imaging to count sorghum panicles directly from plants, demonstrating the effectiveness of computer vision for above-canopy detection.

Beyond these studies, several other works have further advanced grain and spike detection in more complex field conditions. Wen et al. [49] introduced SpikeRetinaNet, an optimized version of RetinaNet tailored for detecting and counting wheat spikes, incorporating a weighted bidirectional feature pyramid network (BiFPN) and focal loss to enhance performance in real-world field environments. Hasan et al. [18] leveraged region-based convolutional neural networks (R-CNNs) trained on the SPIKE dataset to detect and quantify wheat spikes, achieving detection accuracies between 88% and 94% across various growth stages. Zhao et al. [53] further refined the YOLOv5 framework to enable fast and precise identification of wheat grain flaws, demonstrating its potential for real-time quality assessment. For soybean quality evaluation, Momin et al. [31] employed machine vision algorithms based on hue-saturation-intensity (HSI) color models to segment images and detect dockage fractions, achieving defect identification accuracies of up to

98%. Wu et al. [50] also applied deep learning to enumerate wheat grains under varying scenarios and scales, further emphasizing the role of deep learning in yield estimation.

While these studies illustrate the potential of deep learning and computer vision to automate labor-intensive processes and improve grain detection and quality analysis, they predominantly focus on controlled environments or on grains and spikes that remain attached to the plant, where aerial imagery offers a clear visual advantage. In contrast, pre-harvest loss detection introduces unique challenges, including crop shadowing, canopy occlusion, and the presence of visually similar objects (e.g., stones and shades), all of which complicate the reliable identification of fallen grains on the soil surface. These additional complexities highlight the need for specialized approaches tailored to ground-level imaging in post-lodging or pre-harvest contexts.

Our proposed solution uses imaging technologies and deep learning to detect and quantify pre-harvest losses. This approach integrates high-resolution RGB and NIR imaging, and we compared the data quality of aerial and ground data collection platforms. By leveraging a multi-modal approach, combining uncrewed aerial vehicle (UAV) or uncrewed ground vehicle (UGV)-based imaging, the system aims to provide a more comprehensive and scalable alternative to traditional loss monitoring methods, providing more accurate pre-harvest loss estimations. Additionally, by distinguishing between different types of pre-harvest losses — such as natural shattering versus losses caused by mechanical disturbance — this method offers a level of granularity that is not possible with the current quadrat-based technology. When aggregated with post-harvest loss measurements, this segregation of pre-harvest loss sources may lead to more accurate estimations of total machine yield loss and a better understanding of the contributing factors.

The objectives of this study are:

1. Develop a computer vision-based method to replace the traditional quadrat-based approach for measuring pre-harvest loss, enabling automated, scalable, and more precise loss estimation.

2. Introduce the ability to segregate pre-harvest loss sources, providing additional

insights into the origins of grain loss that are currently not captured by existing methods.

3. Evaluate different methods for pre-harvest shedding loss data acquisition.
4. Develop a robust dataset comprising RGB and NIR images of wheat and soybean crops.
5. Implement state-of-the-art deep learning models for the detection and quantification of pre-harvest losses.
6. Validate the models in real-world conditions using an action camera mounted on a harvester.

3.2 Materials and Methods

3.2.1 Platforms and Equipment

Aerial Images

To capture overviews of the crop fields, we utilized a DJI Inspire 2 drone equipped with a Zenmuse camera (DJI-X5, DJI, Shenzhen, Guangdong, China). This setup enabled the collection of a dataset comprising 5145 images of soybean images (5145 images). The images were taken at three different heights— 1.52, 3.05, 4.57, and 6.10 meters above the ground—to ensure diverse perspectives and coverage of the fields. The primary objective of using aerial images was to obtain a broad view of the crop conditions and evaluate the feasibility of detecting pre-harvest loss at different heights.

Handheld Cameras

Handheld cameras were employed to capture both Near-Infrared (NIR) and RGB images. Images were taken at a height of 0.3 m, focusing on the details of grain shedding. This method allowed the collection of high-resolution data essential for identifying minute details not visible in aerial images. The handheld cameras used were XNite Canon ELPH180 (Canon Inc., Tokyo, Japan), with one modified to produce BG-NIR images (680-800 nm).

Rover-based System

A UGV (Husky, Clearpath Robotics, Kitchener, Ontario, Canada) equipped with a stereo camera (RealSense D435i, Intel, Santa Clara, California, USA) adapted to collect ground-level data (Figure 3.1). This rover-based system captured both RGB and IR images at heights of 15 cm and 20 cm. The primary purpose of this setup was to navigate between rows of crops, providing a ground-level view that could reveal additional insights into pre-harvest loss and generate the raw dataset for training a deep learning model for object detection and recognition. The UGV also had an adjustable track width from 0.40 to 0.80 m, allowing it to be adapted to different crop row spacings.

A snout was fabricated to protect the camera and provide artificial light using a 15 cm LED light bar (VEHICODE 6, VEHICODE, California, USA) and an IR Illuminator (850 nm, (Univivi, Hong Kong, China)) (Figure 3.2). Images were also collected at different times of day during summer (Morning, Afternoon, Night) to assess the system's performance under various lighting conditions.

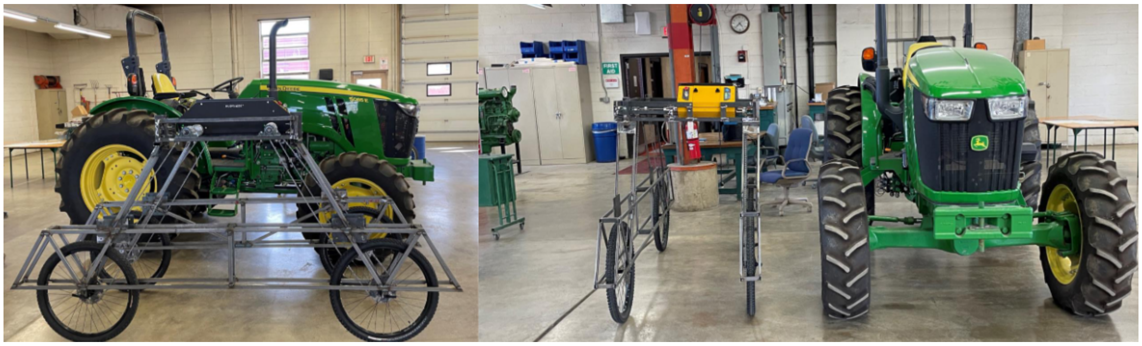


Figure 3.1: Left Image: UGV (front) and a row-crop tractor (Back). Right image: UGV (left) and a row-crop tractor (right). The UGV is equipped with a RealSense D435i camera and fabricated snout for protection and lighting, which is used for ground-level data collection. The John Deere Tractor provides a reference for the scale and utility of agricultural machinery.

Figure 3.2: Fabricated UGV snout equipped with a RealSense D435i camera, a 15 cm LED Light Bar, and an IR Illuminator (850 nm). The snout provides protection for the camera and artificial lighting for improved image quality under various lighting conditions.

Action camera on Harvester

To validate the trained models in real-time, an action camera (Hero 11, GoPro Inc., San Mateo, California, USA) was mounted on a header (Model 3020-30, Case-IH, Racine, Wisconsin, USA) between the two center sickle sections the camera located with a metal bracket orienting it parallel to the direction of travel and 0.15 m above the ground (Figure 3.3). This setup was used to record videos during the harvesting process, which were later used to assess the model's performance in actual field conditions. The video data provided a dynamic evaluation of the model's utility to detect pre-harvest loss.

Figure 3.3: Action camera mounted on the harvester capturing video during the harvesting process. The video data is used to validate the trained models' performance in detecting pre-harvest loss in real-time field conditions.

3.2.2 Data Collection Procedures

Aerial Survey

Aerial images were systematically captured by flying the DJI Inspire 2 drone over the designated soybean fields at heights of 1.52, 3.05, 4.57, and 6.10 meters in October of 2021, on the harvest week at West Madison Agricultural Station (Madison, Wisconsin, USA, 43.061°N, 89.533°W). The flights were conducted in a grid pattern to ensure comprehensive coverage and minimize any gaps in the data having 70% overlap.

Manual Survey

Handheld cameras were used to take close-up images of the soybean crops at a height of 0.30 m in the same time period at Arlington Research Station (Arlington, Wisconsin, USA, 43.303°N, 89.346°W). This process involved manually navigating through the fields and capturing images whenever losses were observed to ensure a representative sample of the objects to be detected.

UGV Survey

The UGV, equipped with a RealSense D435i camera, was deployed to navigate between rows of crops. The images were captured at heights of 15 and 20 cm above the ground, providing a ground-level perspective that complemented the aerial and handheld images. The rover logged images at a rate of 30 frames per second at a resolution of 640 by 480 pixels, along with the robot's odometry data, which allowed us to estimate the navigated distance. This setup enabled the extraction of images from the log at intervals of every 1 meter, avoiding overlapped images.

The UGV was used to collect data for wheat and soybean in the Summer and Fall of 2023, at Arlington Research Station (Arlington, Wisconsin, USA). Data was collected by systematically navigating between crop rows at predetermined intervals, capturing both RGB and IR images at ground-level heights as described previously. The rover's adjustable track width ensured compatibility with varying row spacings, and the onboard lighting system, including LED bars and IR illuminators, enabled consistent image acquisition under diverse ambient lighting conditions. Images were recorded continuously as the UGV traversed the field, and location data (odometry and GPS) facilitated the spatial mapping of collected images. To minimize redundant data and ensure balanced coverage of the study area, images were extracted from the video logs at fixed spatial intervals. Data collection extended from the period immediately before the fields were scheduled for harvest and continued for two subsequent weeks to capture conditions representative of a late harvest scenario. This approach allowed for a comprehensive and high-resolution dataset encompassing different growth stages, lighting scenarios (morning, afternoon, and night), and environmental conditions encountered throughout the summer and fall seasons.

Illumination Survey

We curated a UGV-based dataset of wheat images captured under both natural daylight and artificial illumination conditions during the Summer of 2023. Natural daylight images were recorded at various times throughout the day. In contrast, artificial illumination images were collected by enclosing the camera's field of view within a box, thereby isolating it from ambient light. This setup yielded three distinct datasets designed to assess how daylight and shadows influence model performance:

- **Wheat Noon Dataset:** Acquired around midday under natural daylight supplemented by the artificial lighting system.
- **Wheat Controlled Dataset:** Captured concurrently with the Wheat Noon dataset but with the camera enclosed in a box to block ambient light, resulting in a set of images paired with Wheat Noon.
- **Wheat Night Dataset:** Collected after 6 PM under naturally reduced daylight conditions in combination with artificial illumination.

To assess the influence of environmental conditions on model performance, we will compute per-image performance metrics (True Positives, False Positives, False Negatives, Precision, Recall, and F1-score) for each dataset. After evaluating the normality of these distributions, we will apply appropriate statistical tests: a paired comparison (Wilcoxon signed-rank) between the Noon and Controlled datasets, and independent comparisons (Mann–Whitney U) involving the Night dataset. To control for Type I error across multiple comparisons, we will adjust p-values using the Bonferroni correction.

On-board Survey

For the validation phase, a GoPro camera was mounted on the front of the harvester at 20 cm above the ground. The camera recorded videos during the harvesting process with the combine at a speed of 2.25 km/h at a wheat plot in Arlington Research Station, capturing real-time data that was later used to validate the model's predictions.

3.2.3 Data Annotation

Annotation Tools and Process

The Segment Anything Model (SAM) [24] was utilized for data annotation due to its efficiency in creating standardized masks. Annotators labeled the following:

- **Wheat Crops**: Spikes, grains, and husks (palea, glume, lemma) found on the ground.
- **Soybean Crops**: Grains and pods present on the ground.

Three annotators participated, with one overseeing and verifying the final annotations to ensure consistency and quality. All annotators and overseers have biological systems engineering background. The limited bumber of overseers who review the annotations may create a labeling bias.

Consistency, Quality Control, and Challenges

Consistency was achieved using SAM to create uniform masks and a supervision mechanism to verify all final annotations. The main challenges included ensuring annotation uniformity across multiple annotators and minimizing the impact of natural light variations. To address these, we employed SAM for standardization, designated a supervisor for quality checks, and designed the snout to reduce external light effects. Testing under varying daylight conditions further optimized image quality.

3.2.4 Detection and Localization

This section describes the training process for the models used in detecting and localizing pre-harvest loss. The training was conducted on the following hardware:

- **Processor**: AMD Ryzen 9 7950X 16-Core Processor
- **Operating System**: Ubuntu 20.04.6 LTS
- **Memory**: 64 GB DDR4 RAM

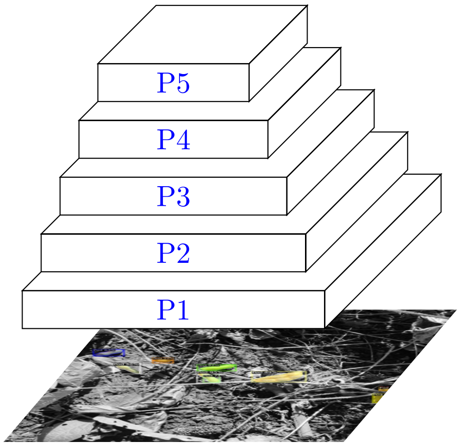


Figure 3.4: YOLOv8 Backbone - custom CSPDarknet53 with cross-stage partial connections.

- **Graphics Card**: NVidia RTX 4090 24GB
- **NVidia Driver**: 520.61.05
- **CUDA Version**: 11.8

The models trained for this study included Mask RCNN [19], YOLOX [15], DETR [8], and a modified YOLOv8-p2, which incorporates the P2 layer detection from YOLOv8 backbone (Figure 3.4) to enhance the detection of smaller objects [13]. These models were selected due to their status as state-of-the-art in object detection, providing robust capabilities for accurately identifying and localizing pre-harvest losses in various crop images. The training process involved fine-tuning each model to optimize performance on the dataset, ensuring that even minor instances of pre-harvest loss were accurately detected. In addition, the NIR channel was included to evaluate if there are improvements to the detection.

Each model was trained using pre-trained models on the COCO dataset [27] (118,000 images) and fine-tuned on our training dataset for 30 epochs. The batch size was set to 1. Data augmentation techniques included saturation adjustments between -25% and +25%, blurring up to 1.5 pixels, and mosaic augmentation. These augmentation strategies were employed to enhance the models' utility to generalize across different lighting and environmental conditions in the dataset. All models re-scaled images to 640×640 prior to inference.

The training process involved fine-tuning each model to optimize performance on the dataset, ensuring that even minor instances of pre-harvest loss were accurately detected. Data were split into 70% for training, 20% for validation, and 10% for testing.

BoT-SORT was utilized to track and count pre-harvest losses in the validation videos obtained from the GoPro [1]. This method was selected for its efficiency and accuracy in object tracking with easy integration with deep learning models for object detection, making it suitable for evaluating the model with the best performance.

3.3 Metrics

In this study, the performance of our models using Precision, Recall, and F1-score was evaluated. These metrics are widely used in object detection and classification tasks to provide a comprehensive assessment of model performance.

3.3.1 Precision

Precision is defined as the ratio of true positive detections to the total number of positive detections (true positives and false positives). It measures the accuracy of the model in identifying only relevant instances.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3.1}$$

3.3.2 Recall

Recall, also known as sensitivity, is the ratio of true positive detections to the total number of actual positives (true positives and false negatives). It evaluates the model's ability to identify all relevant instances.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3.2}$$

3.3.3 F1-score

The F1-score is the harmonic mean of Precision and Recall. It provides a single metric that balances both Precision and Recall, it is particularly useful when there is an uneven class distribution.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.3}$$

3.4 Results and Discussion

3.4.1 Collected Data

The collected data for soybean and wheat datasets are summarized in Tables 3.1 and 3.2, respectively. For soybean, images were captured at various heights using manual and automated navigation methods. The wheat dataset was collected using a modified snout on the UGV, capturing images at two different heights.

Table 3.1: Soybean 2022 Fall Data Collection

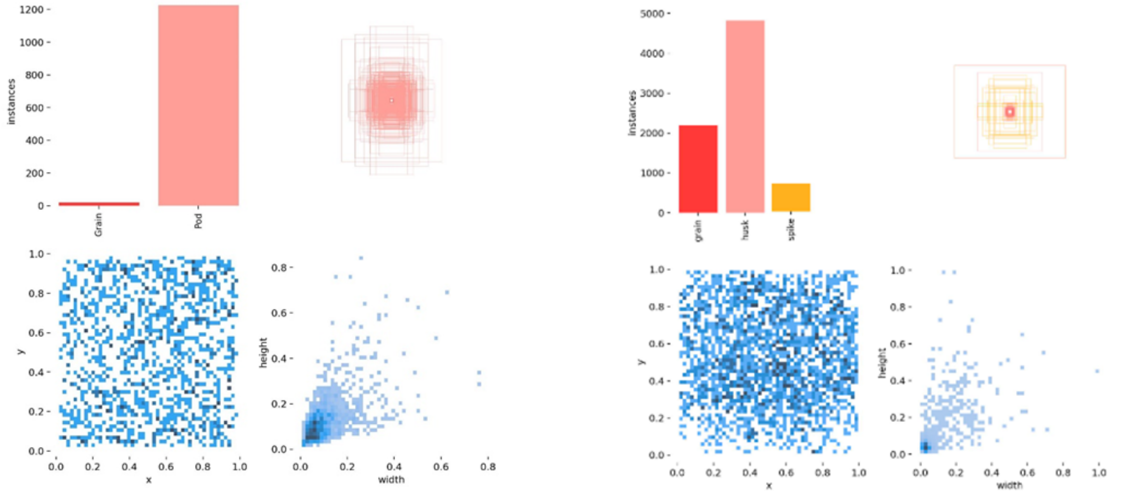
Crop	**Height**	**Navigation**	**NIR Images**	**RGB Images**
Soybean	5 ft	Manual	1784	1136
Soybean	10 ft	Manual	-	75
Soybean	16 ft	Auto	-	1002
Soybean	20 ft	Auto	-	209
Soybean	1 ft	Handheld	400	538

Table 3.2: Wheat Data Collection

Crop	**Mode**	**Navigation**	**NIR Images**	**RGB Images**
Wheat	Snout 6”	UGV	4714	2357
Wheat	Snout 10”	UGV	3746	1873

Figures 3.5a and 3.5b provide visualizations of the datasets. These figures illustrate the distribution of instances, bounding box sizes, and spatial distribution of bounding box centers for the soybean and wheat datasets, respectively. As we can have some false

detections of losses of spikes or pods that are still attached to the plant, we can create filters based on the area of the objects and the location of the bounding boxes, ensuring the detections will occur mainly at middle of the rows. for size distributions, we removed any detection that has more than one standard deviation apart from the mean.



(a) Soybean Dataset. Classes: Grain and Pod.

(b) Wheat Dataset. Classes: Grain, Husk and Spike.

Figure 3.5: Visualization of the datasets. Top-left: Number of instances for each class. Top-right: Distribution of bounding box sizes. Bottom-left: Spatial distribution of bounding box centers (x, y). Bottom-right: Distribution of bounding box widths and heights.

3.4.2 Aerial versus Ground data

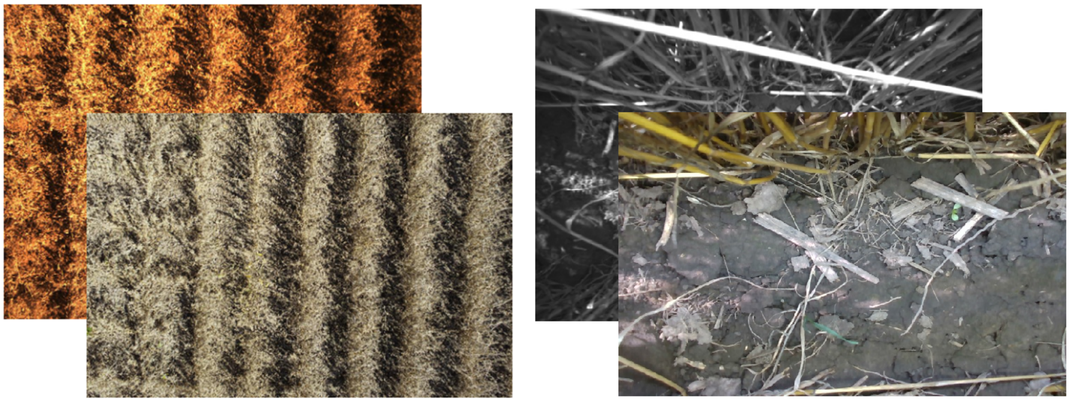


Figure 3.6: Comparison between images acquired by UAV (left) and UGV (right). Back: NDVI and NIR images. Front: RGB images.

The images acquired from the UAV at a 1.5 m altitude (Figure 3.6) lacked sufficient resolution to clearly distinguish grains, debris, or subtle shading variations on the ground. Additionally, the downwash generated by the UAV caused significant disturbance to the crop, potentially dislodging pods or grains. In contrast, the imagery captured by the UGV provided a clearer and more detailed view of the ground, making it possible to annotate losses accurately. Therefore, for pre-harvest loss detection, we opted to use UGV imagery.

3.4.3 Performance Metrics

The performance of the detection models was evaluated using Precision, Recall, and F1-score. Table 3.3 summarizes these metrics for each model on the soybean and wheat datasets.

Table 3.3: Performance Metrics of Models on Soybean and Wheat Datasets

Dataset	**Model**	**Precision**	**Recall**	**F1-score**
Soybean	Mask RCNN	0.676	0.674	0.675
Soybean	Mask RCNN + NIR	0.677	0.692	0.684
Soybean	YOLOX	0.579	0.207	0.305
Soybean	YOLOX + NIR	0.575	0.210	0.307
Soybean	YOLOv8	0.707	0.512	0.594
Soybean	YOLOv8 + NIR	0.723	0.531	0.613
Soybean	YOLOv8-p2	0.727	0.694	**0.710**
Soybean	YOLOv8-p2 + NIR	0.741	0.689	**0.715**
Soybean	DETR	0.299	0.592	0.396
Soybean	DETR + NIR	0.292	0.612	0.396
Wheat	Mask RCNN	0.405	0.315	0.355
Wheat	Mask RCNN + NIR	0.418	0.335	0.372
Wheat	YOLOX	0.337	0.182	0.237
Wheat	YOLOX + NIR	0.327	0.190	0.240
Wheat	YOLOv8	0.697	0.566	0.624
Wheat	YOLOv8 + NIR	0.712	0.584	0.642
Wheat	YOLOv8-p2	0.709	0.688	**0.698**
Wheat	YOLOv8-p2 + NIR	0.729	0.690	**0.709**
Wheat	DETR	0.178	0.165	0.171
Wheat	DETR + NIR	0.191	0.158	0.173

The results indicate that the modified YOLOv8-p2 model outperformed the other models in both the soybean and wheat datasets, achieving the highest F1 scores. This suggests that incorporating the p2 head significantly enhances the detection of smaller objects, especially for the recall, indicating fewer false negatives will occur. The performance of the models did not show a significant improvement when introducing the NIR imagery (850 nm), indicating that this NIR frequency does not provide enough additional information to the RGB images.

Figure 3.7: Detection of soybean pods on the ground using the modified YOLOv8-p2 model. The detected pods are highlighted with bounding boxes and labeled with confidence scores. The color of the bounding boxes indicates different detected instances, aiding in distinguishing between multiple objects in the image.

The integration of BoT-SORT with YOLOv8-p2 for tracking and counting objects improved the results using the validation videos obtained from the on-harvester Go Pro camera. The inference speed of YOLOv8-p2 was 30ms, meaning it is capable to process 33 frames per second. The precision of the header-mounted camera was 80% and recall of 85%, resulting in a F1 score of 82.42%.

Illumination Results

We assembled three distinct datasets, each containing 117 paired RGB and NIR wheat images captured under both natural daylight (Summer 2023) and artificial illumination (Figure 3.8). These datasets, not included in the training process, were used to assess the performance of the YOLOv8-p2+NIR model. As shown in Table 3.4, we examined the models' ability to detect wheat-related objects under varying illumination conditions, thereby providing insights into their robustness and generalization capabilities.

(a) Wheat Noon Dataset image under daylight illumination. Image taken at a noon of Summer 2023.

(b) Wheat Controlled Dataset image under artificial illumination. Image taken at a noon of Summer 2023.

(c) Wheat Night Dataset sample acquired under natural illumination during late harvesting conditions. Image taken after 6 PM of Summer 2023.

Figure 3.8: Representative images from the curated wheat dataset. These examples illustrate various illumination conditions—daylight and artificial lighting—captured during Summer 2023 and late harvesting periods. Such diversity supports robust object detection and recognition tasks.

Table 3.4: Performance Metrics of Models Wheat Dataset for Illumination

Dataset	**Model**	**Precision**	**Recall**	**F1-score**
Wheat Noon	YOLOv8-p2 + NIR	0.560	0.551	0.524
Wheat Controlled	YOLOv8-p2 + NIR	0.628	0.591	0.582
Wheat Night	YOLOv8-p2 + NIR	0.640	0.598	0.594

For each image in each dataset, we determined the True Positives, False Positives, and False Negatives, and subsequently calculated individual Precision, Recall, and F1-scores. A Shapiro-Wilk normality test analysis indicated that the metric distributions were non-normal (p-value< 0.05). To evaluate whether there were statistically significant differences in mean F1-scores across the various illumination conditions, we conducted a Wilcoxon signed-rank test (paired) for the Noon vs. Controlled comparison, and Mann–Whitney U tests (unpaired) for Night vs. Noon and Night vs. Controlled. Due to the multiple comparisons, we applied a Bonferroni correction to adjust the p-values, ensuring a more rigorous control of the overall Type I error rate. The results are summarized in table 3.5.

From the results, a statistical difference was not observed in performance between the dataset taken after 6 PM with presence of natural light and artificial light and the dataset taken using a controlled environment with only artificial lights. However, daylight (as in the Noon dataset) had a statistically significant impact on precision, suggesting that the model may generate more false positives under bright and shadow-heavy conditions, which in turn affects its F1-score. The recall does not provide a statistical difference, meaning that the number of false negatives are not statistically different between all datasets.

3.4.4 Discussion

The results demonstrate that the modified YOLOv8-p2 model provides better performance in detecting pre-harvest losses in small grain crops compared to other state-of-the-art models. The utility to accurately detect and quantify these losses can significantly aid in understanding the impact the harvester has on field losses. The proposed methodology of using on-harvester imaging and deep learning models presents a scalable and reliable

Table 3.5: Statistical test results for Precision, Recall, and F1-score across Wheat datasets (Noon, Controlled, Night). Normality p-values (Shapiro–Wilk) are provided for each dataset, followed by adjusted p-values (Bonferroni correction) for the paired (Wilcoxon) and independent (Mann–Whitney U) tests.

	Normality p-values			**Adjusted p-values (Bonferroni)**		
Metric	**Noon**	**Controlled**	**Night**	Noon vs Controlled	Night vs Noon	Night vs Controlled
Precision	0.0289	0.0199	0.0128	0.0030	0.0033	0.9981
Recall	0.0084	0.0418	0.0391	0.1213	0.2239	0.8201
F1-Score	0.0447	0.0082	0.0066	0.0011	0.0023	0.9054

solution for pre-harvest loss detection. Future work may focus on refining these models further and exploring their applicability to other crop types.

These findings are consistent with previous studies. Feng et al. (2024) [13] similarly reported that merging the P2 and P3 heads enhanced the detection of small objects using the YOLOv8 architecture on satellite imagery. Likewise, Badgujar et al. (2024) [5], through a bibliometric and systematic review of YOLO-based object detection in agricultural contexts, identified YOLO as one of the most successful models for this task, though it often struggles when objects are small, densely clustered, or affected by dynamic environmental factors such as shadows [6, 34]. Our results confirm that the P2 head bolsters the detection of small objects in crop scenarios, yet emphasize the persistent challenge of illumination. Adequate lighting, whether achieved by timing image capture or using artificial sources, is pivotal for achieving more reliable detection outcomes.

The Vision Transformer (ViT) did not outperform the convolutional neural network (CNN) models, likely due to the limited size of our training dataset compared to the larger datasets typically used for transformer-based architectures. While our results indicate that YOLOv8-p2 achieved the highest performance, it is noteworthy that DETR operates at 86 Giga Floating Operations per Second (GFLOPS) with a frame rate of 28 FPS and has 41 million trainable parameters, whereas YOLOv8-p2 requires 260.6 GFLOPS, operates at 130 FPS, and includes 66.6 million trainable parameters. Given hardware constraints, DETR may still present a viable alternative due to its relatively lower computational requirements.

The high computational demands of deep learning models and the reliance on high-resolution imaging present challenges for real-world deployment, particularly in low-resource or edge computing environments. While we successfully implemented YOLOv8 on a Raspberry Pi 3 (Raspberry Pi Foundation, Cambridge, UK) CPU achieving approximately 2 FPS and optimized inference using the OAK-1 (Luxonis, CO, USA) camera's Myriad VPU, achieving approximately 3 FPS, these solutions still require hardware with sufficient processing capabilities. Additionally, the need for high-resolution cameras may increase equipment costs, limiting accessibility for small-scale farmers compared to traditional manual pre-harvest loss estimation methods. Future work should explore cost-effective alternatives, such as model compression, lower-resolution imaging, or cloud-based inference, to enhance financial feasibility and enable broader adoption of automated pre-harvest loss detection.

Integrating a tracking algorithm such as BoT-SORT leverages the redundancy of sequential frames in a video, each potentially containing the same object. This redundancy ensures that objects missed in one frame may still be detected in subsequent frames, thus reducing false negatives and improving overall detection and counting accuracy. BoT-SORT's robust association mechanism also preserves object identities across frames, providing more consistent data for subsequent analysis. In addition, it lowers computational overhead by efficiently linking detections across frames, making it practical for real-time applications. By using BoT-SORT in our study, we comprehensively evaluated tracked instances of crop losses under real-world conditions, thereby validating the method's effectiveness.

Similar enhancements in object detection and tracking have been reported for a range of agricultural applications, including grapes in vineyards [42], sweetpotatoes [51], chickens [45], and pigs [47]. Saraceni et al. (2024) [42] demonstrated that tracking algorithms significantly improve grape detection in robotic systems, while Xu et al. (2024) [51] and Tu et al. (2024) [47] showed that tracking strategies enhance robustness under varying illumination conditions by detecting objects in different frames and perspectives. Our results align with these findings, indicating that tracking likewise increases the reliability of detecting very small objects.

The experimental settings, which varied light conditions and camera heights, provided insights into optimizing image quality for measuring pre-harvest loss:

- **Camera height:** For a camera with a focal length of 1.93 mm, the optimal height is between 15 to 25 cm above the ground.
- **Exposure:** For on-board cameras, a high shutter speed of 90 fps is recommended to capture sharp images during motion.
- **Controlled illumination:** Using diffuse lighting or a shaded environment can reduce harsh shadows and ensure consistent image quality under varying outdoor conditions.

Our results showed that the near-infrared (NIR, 850 nm) channel did not show significant improvements for detection models at a speed of 2.4 km/h and wavelength of 850. However, there remains a need to validate the use of multispectral and hyperspectral imaging. Combining different wavelengths can yield multiple indices, such as the Normalized Difference Vegetation Index (NDVI) and Soil-Adjusted Vegetation Index (SAVI), which can assist in segmenting grains from the background. Additionally, NIR reflectance characteristics can provide insights into the quality of grain losses [11].

From Figure 3.5a, it appears that under the specific harvest conditions observed in this study, the soybean crop experienced minimal pre-harvest shattering losses. However, post-harvest observations revealed a significant number of grains on the ground, suggesting that under these conditions, the majority of losses were attributable to the harvesting machinery.

By integrating pre-harvest and post-harvest loss measurements, it becomes possible to isolate the specific contributions of harvesting machinery to overall crop loss, thereby enabling targeted improvements in harvesting technology and practices. This integrated approach is crucial for enhancing agricultural efficiency and sustainability.

3.5 Conclusion

The results demonstrated that the modified YOLOv8-p2 model, which incorporates the p2 head for enhanced detection of smaller objects, outperformed other models in both soybean and wheat datasets. This model achieved the highest Precision, Recall, and F1-scores, highlighting its effectiveness in identifying pre-harvest losses. On more restricted hardware scenarios, DETR could still be an option with the use of Shifted Patch Tokenization (SPT) and Locality Self-Attention (LSA), that can improve the results of the model on smaller datasets [26]. This study did not aim to optimize the models to embedded computers, therefore further trade-off analysis from optimizing, pruning and quantizing the models to run on constrained devices would necessary.

The NIR channel provided very low performance improvement. This indicates that the selected wavelength of 850 nm did not provide more information than using only the RGB channels. The limited NIR wavelength may not have captured significant spectral differences between target and background materials, suggesting that either a broader NIR range or a different wavelength selection could enhance feature discrimination. Additionally, the low impact of the NIR channel may be due to the specific crop and environmental conditions during data collection, where the 850 nm reflectance did not introduce sufficient contrast for improved classification.

This study also addressed the limitations of traditional methods for measuring pre-harvest losses, such as the use of quadrats, which are labor-intensive and sparse in coverage. The proposed methodology offers a scalable, efficient, and accurate alternative, minimizing disturbances to the crop while collecting comprehensive data. By combining pre-harvest and post-harvest loss measurements, it is possible to identify inefficiencies and areas for improvement in harvesting technology, ultimately leading to reduced losses and increased agricultural efficiency.

Future work should focus on further refining these models and exploring their applicability to other crop types. Additionally, the integration of real-time processing and automation in data collection and analysis can further enhance the practicality and efficiency of this approach.

References

[1] D. Aharon, M. Benavides, and D. Bashan. “BoT SORT: Robust Associations Multi-Pedestrian Tracking”. In: *arXiv preprint arXiv:2212.09451* (2022).

[2] G. Albahri et al. “Enhancing essential grains yield for sustainable food security and bio-safe agriculture through latest innovative approaches”. en. In: *Agronomy (Basel)* 13.7 (June 2023), p. 1709.

[3] S. Arslan and T. S. Colvin. “Grain Yield Mapping: Yield Sensing, Yield Reconstruction, and Errors”. en. In: *Precis. Agric.* 3.2 (June 2002), pp. 135–154.

[4] R. Avtar and T. Watanabe, eds. *Unmanned aerial vehicle: Applications in agriculture and environment*. en. 1st ed. Cham, Switzerland: Springer Nature, Nov. 2019.

[5] C. M. Badgujar, A. Poulose, and H. Gan. “Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review”. In: *Comput. Electron. Agric.* 223 (Aug. 2024), p. 109090.

[6] H. C. Bazame et al. “Detection, classification, and mapping of coffee fruits during harvest with computer vision”. en. In: *Comput. Electron. Agric.* 183.106066 (Apr. 2021), p. 106066.

[8] N. Carion et al. “End-to-End Object Detection with Transformers”. In: *European Conference on Computer Vision* (2020).

[9] J. del Cerro et al. “Unmanned aerial vehicles in agriculture: A survey”. en. In: *Agronomy (Basel)* 11.2 (Jan. 2021), p. 203.

[10] D. Choi et al. "*A performance comparison of RGB, NIR, and depth images in immature citrus detection using deep learning algorithms for yield prediction*". en. In: *2017 Spokane, Washington July 16 - July 19, 2017*. St. Joseph, MI: American Society of Agricultural and Biological Engineers, 2017, p. 1.

[11] D. R. Massie and K. H. Norris. "Spectral reflectance and transmittance properties of grain in the visible and near infrared". en. In: *Trans. ASAE* 8.4 (1965), pp. 0598–0600.

[12] S. Dong et al. "Automatic crop pest detection oriented multiscale feature fusion approach". en. In: *Insects* 13.6 (June 2022), p. 554.

[13] F. Feng et al. "Improved YOLOv8 algorithms for small object detection in aerial imagery". In: *Journal of King Saud University - Computer and Information Sciences* 36.6 (July 2024), p. 102113.

[14] A. Galieni et al. "Past and future of plant stress detection: An overview from remote sensing to positron emission tomography". en. In: *Front. Plant Sci.* 11 (2020), p. 609155.

[15] Z. Ge et al. "YOLOX: Exceeding YOLO Series in 2021". In: *arXiv preprint arXiv:2107.08430* (2021).

[16] Q. Geng et al. "A rapid, low-cost wheat spike grain segmentation and counting system based on deep learning and image processing". en. In: *Eur. J. Agron.* 156.127158 (May 2024), p. 127158.

[17] T. Guo, T. Kujirai, and T. Watanabe. "MAPPING CROP STATUS FROM AN UNMANNED AERIAL VEHICLE FOR PRECISION AGRICULTURE APPLICATIONS". en. In: XXXIX-B1 (July 2012), pp. 485–490.

[18] M. M. Hasan et al. "Detection and analysis of wheat spikes using Convolutional Neural Networks". en. In: *Plant Methods* 14.1 (Nov. 2018), p. 100.

[19] K. He et al. "Mask R-CNN". In: *Proceedings of the IEEE international conference on computer vision* (2017).

[20] L. Hume. “Crop Losses in Wheat (Triticum aestivum) as Determined Using Weeded and Nonweeded Quadrats”. In: *Weed Sci.* 33.5 (1985), pp. 734–740.

[21] A. Hussain et al. *Automatic disease detection in wheat crop using convolution neural network.* 2018.

[22] C. James et al. “GrainPointNet: A deep-learning framework for non-invasive sorghum panicle grain count phenotyping”. en. In: *Comput. Electron. Agric.* 217.108485 (Feb. 2024), p. 108485.

[23] B. Jiang et al. “MGLNN: Semi-supervised learning via multiple graph cooperative Learning Neural Networks”. en. In: *Neural Netw.* 153 (Sept. 2022), pp. 204–214.

[24] A. Kirillov et al. “Segment Anything”. In: *ICCV* (Apr. 2023), pp. 3992–4003.

[25] O. Kulkarni. “Crop disease detection using deep learning”. en. In: *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).* IEEE, Aug. 2018, pp. 1–4.

[26] S. H. Lee, S. Lee, and B. C. Song. “Vision Transformer for Small-Size Datasets”. In: *arXiv [cs.CV]* (Dec. 2021).

[27] T.-Y. Lin et al. “Microsoft COCO: Common objects in context”. In: *European conference on computer vision* (2014), pp. 740–755.

[28] Z. Lin and W. Guo. “Sorghum panicle detection and counting using unmanned aerial system images and deep learning”. en. In: *Front. Plant Sci.* 11 (Sept. 2020), p. 534853.

[29] Z. Liu et al. “Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion”. en. In: *IEEE Access* 8 (2020), pp. 2327–2336.

[30] Á. Mesterházy, J. Oláh, and J. Popp. “Losses in the Grain Supply Chain: Causes and Solutions”. en. In: *Sustain. Sci. Pract. Policy* 12.6 (Mar. 2020), p. 2342.

[31] M. A. Momin et al. “Machine vision based soybean quality evaluation”. In: *Comput. Electron. Agric.* 140 (Aug. 2017), pp. 452–460.

[32] N. S. Monhollen et al. “In-field machine vision system for identifying corn kernel losses”. In: *Comput. Electron. Agric.* 174 (July 2020), p. 105496.

[33] L. Mosley et al. “Image-based sorghum head counting when you only look once”. In: *arXiv [cs.CV]* (Sept. 2020).

[34] A. Nasirahmadi, U. Wilczek, and O. Hensel. “Sugar beet damage detection during harvesting using different convolutional neural network models”. en. In: *Agriculture* 11.11 (Nov. 2021), p. 1111.

[35] C. Y. N. Norasma et al. “Unmanned aerial vehicle applications in agriculture”. en. In: *IOP Conf. Ser. Mater. Sci. Eng.* 506.1 (Apr. 2019), p. 012063.

[36] M. Ouhami et al. “Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research”. en. In: *Remote Sens. (Basel)* 13.13 (June 2021), p. 2486.

[37] M. Pineda, M. Barón, and M.-L. Pérez-Bueno. “Thermal imaging for plant stress detection and phenotyping”. en. In: *Remote Sens. (Basel)* 13.1 (Dec. 2020), p. 68.

[38] M. F. F. Rahman et al. “A comparative study on application of unmanned aerial vehicle systems in agriculture”. en. In: *Agriculture* 11.1 (Jan. 2021), p. 22.

[39] A. M. Roy and J. Bhaduri. “DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism”. en. In: *Adv. Eng. Inform.* 56.102007 (Apr. 2023), p. 102007.

[40] A. M. Roy and J. Bhaduri. “Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4”. en. In: *Comput. Electron. Agric.* 193.106694 (Feb. 2022), p. 106694.

[41] A. M. Roy, R. Bose, and J. Bhaduri. “A fast accurate fine-grain object detection model based on YOLOv4 deep neural network”. en. In: *Neural Comput. Appl.* 34.5 (Mar. 2022), pp. 3895–3921.

[42] L. Saraceni et al. “AgriSORT: A Simple Online Real-time Tracking-by-Detection framework for robotics in precision agriculture”. en. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. Vol. 2. IEEE, May 2024, pp. 2675–2682.

[43] U. Shafi et al. “Wheat rust disease detection techniques: a technical perspective”. en. In: *J. Plant Dis. Prot. (2006)* 129.3 (June 2022), pp. 489–504.

[44] R. Shukla et al. “Detecting crop health using machine learning techniques in smart agriculture system”. en. In: *J. Sci. Ind. Res. (India)* 80.08 (Aug. 2021), pp. 699–706.

[45] A. L. R. Siriani et al. “Chicken tracking and individual bird activity monitoring using the BoT-SORT algorithm”. en. In: *AgriEngineering* 5.4 (Sept. 2023), pp. 1677–1693.

[46] J. Sun et al. “Accurate rice grain counting in natural morphology: A method based on image classification and object detection”. en. In: *Comput. Electron. Agric.* 227.109490 (Dec. 2024), p. 109490.

[47] S. Tu et al. “Tracking and automatic behavioral analysis of group-housed pigs based on YOLOX+BoT-SORT-slim”. en. In: *Smart Agricultural Technology* 9.100566 (Dec. 2024), p. 100566.

[48] R. Wang et al. “S-RPN: Sampling-balanced region proposal network for small crop pest detection”. en. In: *Comput. Electron. Agric.* 187.106290 (Aug. 2021), p. 106290.

[49] C. Wen et al. “Wheat spike detection and counting in the field based on SpikeRetinaNet”. en. In: *Front. Plant Sci.* 13 (Mar. 2022), p. 821717.

[50] W. Wu et al. “Detection and enumeration of wheat grains based on a deep learning method under various scenarios and scales”. en. In: *J. Integr. Agric.* 19.8 (Aug. 2020), pp. 1998–2008.

[51] J. Xu and Y. Lu. “Prototyping and evaluation of a novel machine vision system for real-time, automated quality grading of sweetpotatoes”. en. In: *Comput. Electron. Agric.* 219.108826 (Apr. 2024), p. 108826.

[52] J. Zhang et al. “Chlorophyll content detection of field maize using RGB-NIR camera”. en. In: *IFAC-PapersOnLine* 51.17 (Jan. 2018), pp. 700–705.

[53] W. Zhao et al. “Fast and accurate wheat grain quality detection based on improved YOLOv5”. en. In: *Comput. Electron. Agric.* 202.107426 (Nov. 2022), p. 107426.

[54] H. Zheng et al. “Early season detection of rice plants using RGB, NIR-G-B and multispectral images from unmanned aerial vehicle (UAV)”. en. In: *Comput. Electron. Agric.* 169.105223 (Feb. 2020), p. 105223.

Chapter 4

Integrating RGB-Derived Morphometrics and NIR Spectra for Robust Optical Estimation of Forage Processing Quality

William Yamada, Anmol Sharma, Tess Lemcke, Kevin Shinners, Benedikt Jung, Troy Runge and Matthew Digman

Publication Note: This chapter presents original research that is *intended for submission to Computers and Electronics in Agriculture.* The author of this dissertation led the experimental design, multimodal data fusion methodology, statistical analysis, and manuscript preparation.

Abstract: Accurate assessment of Corn Silage Processing Score (CSPS) is critical for evaluating forage quality and optimizing livestock nutrition. Traditional laboratory methods, though reliable, are time-consuming, destructive, and incompatible with real-time field applications. This study proposes a multimodal machine learning framework to predict CSPS using both RGB image-derived particle metrics and near-infrared (NIR) spectral features. A dataset comprising laboratory CSPS measurements, image-based kernel fragmentation statistics, and on-harvester NIR estimates was collected from 54 silage plots under varying harvester settings. We evaluated unimodal and multimodal fusion strategies, including early and late fusion, proxy modeling using the Fragment of Grain (FoG) approach, and various regression techniques such as Random Forest, XGBoost, Gaussian Process Regression, and Bayesian Neural Networks.

Results show that RGB-based models perform best when predicting FoG as an intermediate variable ($R^2 = 0.727$, RMSE = 2.61%), while NIR models are more effective for direct CSPS estimation ($R^2 = 0.911$, RMSE = 2.69%). Although early fusion underperformed relative to unimodal models, late fusion strategies that combine RGB and NIR predictions yielded the highest accuracy ($R^2 = 0.949$, RMSE = 1.68%) using Random Forest regression. These findings demonstrate the potential of integrating multimodal sensing and interpretable machine learning for scalable, non-invasive evaluation of forage processing quality in precision agriculture.

4.1 Introduction

Corn silage is a foundational component of livestock rations, supplying both fiber and a major source of energy through starch. The degree to which starch is available to ruminants depends heavily on the effectiveness of kernel processing during harvest. To quantify this, the Corn Silage Processing Score (CSPS) has been established as a key indicator, defined as the proportion of starch passing through a 4.75 mm sieve [10]. This measurement captures the extent of kernel fragmentation, which influences both starch digestibility and subsequent animal performance [9].

Despite its importance, CSPS assessment is constrained by its dependence on post-harvest laboratory methods, including sieving [27] and wet chemistry protocols [10]. These approaches are time-consuming, labor-intensive, and inherently limited in their ability to support in-field, real-time adjustments to harvester settings.

4.1.1 Related Works

The initial attempts to predict CSPS from images focused on traditional computer vision techniques. Drewry et al. [7] separated corn silage particles from the background and used contour-based features to estimate kernel fragmentation, reporting high correlations with laboratory CSPS, particularly when the number of large particles was low. Building on that, Luck et al. [20] applied image analysis to predict the in situ dry matter disappearance of chopped kernels, linking visual metrics with digestibility estimates.

More recent approaches have shifted toward data-driven methods using machine learning. Rocha et al. [7] implemented real-time RGB imaging on harvesters to assess CSPS through kernel detection and regression modeling. Rasmussen and Moeslund [24, 23] employed deep convolutional networks, such as R-FCN, to identify kernel fragments directly in non-separated silage samples. However, their work acknowledged the challenges of precise annotation and stressed the need for robust evaluation beyond bounding-box metrics. This concern was further explored in [22], where they highlighted how annotation variability and occlusion affect the reliability of deep learning models, especially when comparing optical estimates to reference CSPS values.

This work is the first to systematically compare early and late fusion strategies that integrate RGB-derived morphological features and near-infrared (NIR) spectral features for the prediction of Corn Silage Processing Score (CSPS). By evaluating both proxy-based models—using the Fragment of Grain (FoG) as an intermediate estimator—and direct prediction approaches, we highlight the relative strengths of each sensing modality and modeling pipeline. Moreover, this study introduces a comprehensive analysis of feature selection methods, dimensionality reduction techniques, and regression model calibration in the context of whole-silage image datasets—an area previously unexplored in the literature. Importantly, we also examine the correlation between predicted and laboratory-reference CSPS values and quantify model uncertainty through cross-validation. These contributions lay the groundwork for robust, field-deployable, multimodal systems for non-destructive silage quality assessment.

4.1.2 Objectives

The primary objective of this study was to develop and evaluate models for optical CSPS estimation using image-derived features and spectral data. Specifically, we aim to:

- Quantify the predictive power of RGB image features that reflect particle size, shape, and spatial distribution;
- Evaluate the added benefit of near-infrared (NIR) spectroscopy for capturing compositional properties associated with kernel breakage;
- Compare early fusion (joint feature modeling) and late fusion (independent modeling with combined outputs) strategies for multimodal prediction;
- Investigate signal-to-noise ratios and inter-laboratory variability to understand fundamental limitations in CSPS prediction.

This work builds on prior image-based and sensor-based methods but uniquely combines multiple data sources under a unified framework, advancing the feasibility of scalable, real-time assessment of forage processing effectiveness.

4.2 Materials and Methods

4.2.1 Image collection

The image dataset was acquired during the 2021 harvest season in Arlington, Wisconsin. Sampling was conducted on 30 plots across two distinct dates: early and late harvest, with a five-day interval between them. To simulate real-world variability, the forage harvester's machine settings were randomized, encompassing three Theoretical Length of Cut (TLOC) levels (7, 19, and 25 mm), three hybrids, and three Roll Gap levels (1, 2, and 3 mm).

For each plot, a minimum of 30 high-speed images was captured as the corn silage exited the spout of the forage harvester. These images were obtained using a John Deere FQ Camera (4.1), known for its capability to deliver high-resolution imaging at rapid frame rates, which is critical for capturing the dynamic flow of silage particles. This approach provided a detailed visual dataset reflective of diverse field conditions.

Figure 4.1: Setup of sensing devices on the forage harvester spout. The green device is the FQ camera, which was used to capture high-speed RGB imagery of corn silage material during harvesting for optical feature extraction. The yellow box corresponds to the HarvestLab near-infrared (NIR) sensor, which performs real-time chemical composition estimations such as starch, crude protein, and fiber contents. Both systems were installed upstream of the material discharge, allowing synchronized acquisition of physical and compositional properties for Corn Silage Processing Score (CSPS) estimation.

To prepare the images for further analysis, each one was cropped to isolate the region of interest (ROI) by segmenting the region that is within the glass view, generating images of size 1836 px × 468 px. This preprocessing step enhanced the precision of subsequent data annotation and feature extraction, ensuring the dataset's reliability for modeling and statistical analysis.

4.2.2 Data annotation

Each image underwent annotation using the Segment Anything Model (SAM) [17], a segmentation tool designed for efficient object annotation. Five trained labelers manually annotated images to ensure consistent segmentation of silage particles. Each annotator followed a standardized training process, and two reviewers validated the annotations for quality control. All annotators had agricultural or biological systems engineering backgrounds. Only particles that the annotators were confident was a kernel fragment were annotated, thus leaving some some kernel particles unannotated.

A total of 30 images from each of the 60 plots (30 plots × 2 collection dates) were annotated. For each annotated image, the following metrics were extracted:

- Maximum Inscribed Diameter [20, 7]: The diameter of the largest circle that can fit within the segmented particle, providing a measure of particle size.
- Major and Minor Axes [24, 23]: The lengths of the primary and secondary axes of an ellipse fitted to the segmented particle, capturing its overall shape.
- Area: The total number of pixels within the segmented region, representing the particle's projected area.

In addition to these primary metrics, statistical measures were computed for a more granular analysis:

- Quantiles: The 25th, 50th (median), and 75th percentiles for each metric, denoted as $Metric_{Q25}$, $Metric_{Q50}$, and $Metric_{Q75}$ (e.g., $Area_{Q25}$, $Area_{Q50}$, $Area_{Q75}$).
- Counts of Kernels Below Quantiles: The number of particles smaller than each quantile, calculated for each metric (e.g., $Count_{Area<Q25}$, $Count_{Area<Q50}$, $Count_{Area<Q75}$).
- Relative Area Under Quantiles: The proportion of the total segmented area occupied by particles below each quantile (e.g., $Area_{Metric<Qxx}$), expressed as a ratio (e.g., $Area_{Area<Q25}/TotalArea$).

Table 4.1: Statistical summary of the Corn Silage Processing Score (CSPS) values obtained from two laboratories and their average. The table includes the number of samples (N), mean, standard deviation, minimum, median, and maximum CSPS values, highlighting consistency between the two laboratory measurements and their aggregated average.

Source	N	Mean	Standard Deviation	Minimum	Median	Maximum
Laboratory 1	60	74.68	8.13	56.67	75.77	87.57
Laboratory 2	60	75.81	8.29	52.90	76.98	88.42
Average	60	75.24	8.18	54.78	76.53	88.09

4.2.3 On-Board Near Infrared Spectrometer and Machine Settings Data

During the harvest, we utilized a HarvestLab spectrometer (Model 3000, John Deere) to measure the near-infrared (NIR) spectrum (950 - 1650 nm, 256 diodes) of corn silage as it exited the spout. At least 5 spectra were obtained from each plot and averaged. The system enabled the collection of on-board estimations for key compositional attributes of the corn silage, including Moisture Content, Starch, Protein, Neutral Detergent Fiber (NDF), Acid Detergent Fiber (ADF), and Ash. Along with the NIR quality estimations, we also annotated the machine's theoretical length-of-cut (TLOC) and the roll gap settings for each plot.

4.2.4 Algorithm Development

Optical CSPS Estimation

To derive the Corn Silage Processing Score (CSPS) from images, we estimated it based on the area-weighted cumulative undersize percent of 4.75 mm, a threshold commonly used in particle size analysis. While several studies have investigated optical methods for particle size measurement, different metrics have been proposed in the literature to quantify kernel size and fragmentation.

For defining the size threshold of particles under 4.75 mm, we measured:

- Maximum Inscribed Circle Diameter (MICD)
- Rotated Bounding Box Large Axis

- Rotated Bounding Box Small Axis
- Bounding Ellipsoid Major Axis
- Bounding Ellipsoid Minor Axis
- Area compared to the area of a circle with 4.75 mm diameter (17.72 mm^2)

For estimating starch content, we evaluated:

- Segmented Area
- Bounding Ellipsoid Volume
- Enhanced Volume

The bounding ellipsoid volume method approximates kernels as 3D ellipsoids, leveraging the geometric properties of ellipsoids to estimate volume. This approach uses the formula:

$$V = \pi \cdot \frac{a \cdot b^2}{6} \tag{4.1}$$

where a represents the major axis and b the minor axis of the fitted ellipse. This method assumes that the depth of the kernel is equivalent to the minor axis, which simplifies the calculation but may not accurately capture the true volumetric characteristics of the kernel, especially when the kernel's shape deviates from a perfect ellipsoid.

In contrast, the enhanced volume method refines this approximation by incorporating the MICD to estimate the depth dimension. The formula used is:

$$V = \pi \cdot \frac{a \cdot b \cdot r_{MICD}}{3} \tag{4.2}$$

where r_{MICD} is the diameter of the largest circle that can be inscribed within the kernel's cross-section. This approach provides a more nuanced representation of kernel thickness, accounting for variations in kernel shape that the bounding ellipsoid method might overlook. By using MICD, the enhanced volume method captures additional structural information, leading to improved volumetric estimation and a closer alignment with the actual physical properties of the kernels. This refinement is particularly beneficial in scenarios where kernels are partially visible or irregularly shaped, as it mitigates the limitations of assuming uniform depth based solely on the minor axis.

Since kernels can be partially visible due to obstructions in the images, their measured dimensions may not always reflect their true size. To account for this potential bias, we also calculated CSPS using different thresholds beyond the standard 4.75 mm, considering the 25th, 50th, and 75th quantiles of the kernel size distribution. This approach allowed us to assess the impact of partial kernel views on CSPS estimation and determine whether alternative thresholding strategies could improve the robustness of optical measurements.

This multi-metric approach ensured a comprehensive evaluation of optical CSPS estimation, capturing both particle size and starch content while addressing challenges related to kernel visibility and obstruction in the imaging process.

Figure 4.2 illustrates the pipeline for RGB-based optical CSPS estimation. The figure begins with a raw corn silage image from which kernel segmentation was completed. Individual kernels are identified using image segmentation techniques and then analyzed to extract a set of morphological descriptors. These include shape (e.g., major/minor axis lengths), size (e.g., area, volume), and distribution statistics (e.g., counts of large fragments exceeding a specific quantile). These descriptors are then used in two branches: (1) to calculate the Fragment of Grain (FoG) score, which serves as a proxy for CSPS, and (2) as input features in machine learning models that directly estimate CSPS. This dual pathway highlights how RGB imagery provides both indirect and direct routes to infer kernel processing quality.

Feature Selection

Feature selection plays a critical role in developing accurate and interpretable predictive models, particularly when combining image-derived metrics and spectral data. Identifying the most relevant predictors reduces the risk of overfitting, improves interpretability, and

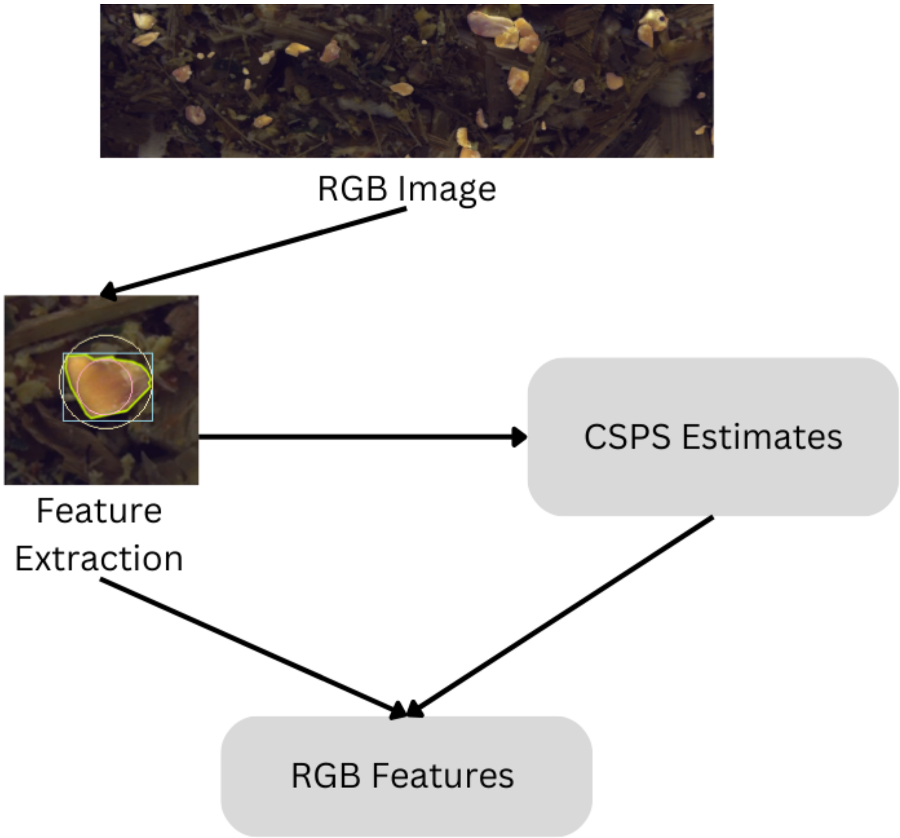


Figure 4.2: Schematic representation of the optical CSPS estimation process. Features extracted from RGB images are used to derive CSPS estimates using the morphological features from the corn kernels.

reduces computational complexity.

Given the established importance of Partial Least Squares (PLS) regression in near-infrared (NIR) spectroscopy analysis, PLS served as the baseline method for selecting relevant wavelengths from the average spectrum of each plot. The data was preprocessed using Standard Normal Variate (SNV) followed by a Savitzky-Golay filter with a window size of 11, polynomial order of 2, and first derivative. In addition to PLS, we applied complementary wavelength selection techniques, including:

- Partial Least Squares Regression - Variable Importance in Projection (PLSR-VIP) [21, 8], using a 5-fold cross-validation and a consensus selection where wavelengths were selected if at least 3 of the 5 folds were agreed on.

- Competitive Adaptive Reweighted Sampling (CARS) [14, 13].
- Successive Projections Algorithm (SPA) [28, 15].
- Uninformative Variable Elimination (UVE) [18, 15].
- Stepwise Selection (SS) [33, 16].
- Principal Component Analysis (PCA) [11, 19].

These wavelength selection methods were applied to the average NIR spectrum per plot, ensuring that only the most relevant spectral features were included in the predictive models. Figure 4.3 presents the NIR-based feature extraction and selection process. It starts with the average NIR reflectance spectra acquired from each silage plot. The raw spectra are subjected to preprocessing steps including Standard Normal Variate (SNV) correction and Savitzky-Golay filtering to reduce baseline shifts and noise. The cleaned spectra were then passed through various wavelength selection algorithms—CARS, UVE, SPA, VIP, and stepwise selection (SS)—to isolate the most informative spectral bands. The selected wavelengths served as predictors in machine learning models trained to estimate CSPS.

To further assess the relative importance of all features, including image metrics, spectral features, and machine parameters (TLOC, roll gap), we applied the following techniques, as described in [4, 30]:

- Pearson correlation, to capture linear relationships with CSPS.
- Spearman correlation, to capture monotonic but potentially nonlinear relationships.
- Mutual Information (MI), to quantify both linear and nonlinear dependency between each feature and CSPS.
- Recursive Feature Elimination (RFE), to iteratively remove less relevant features based on model performance.

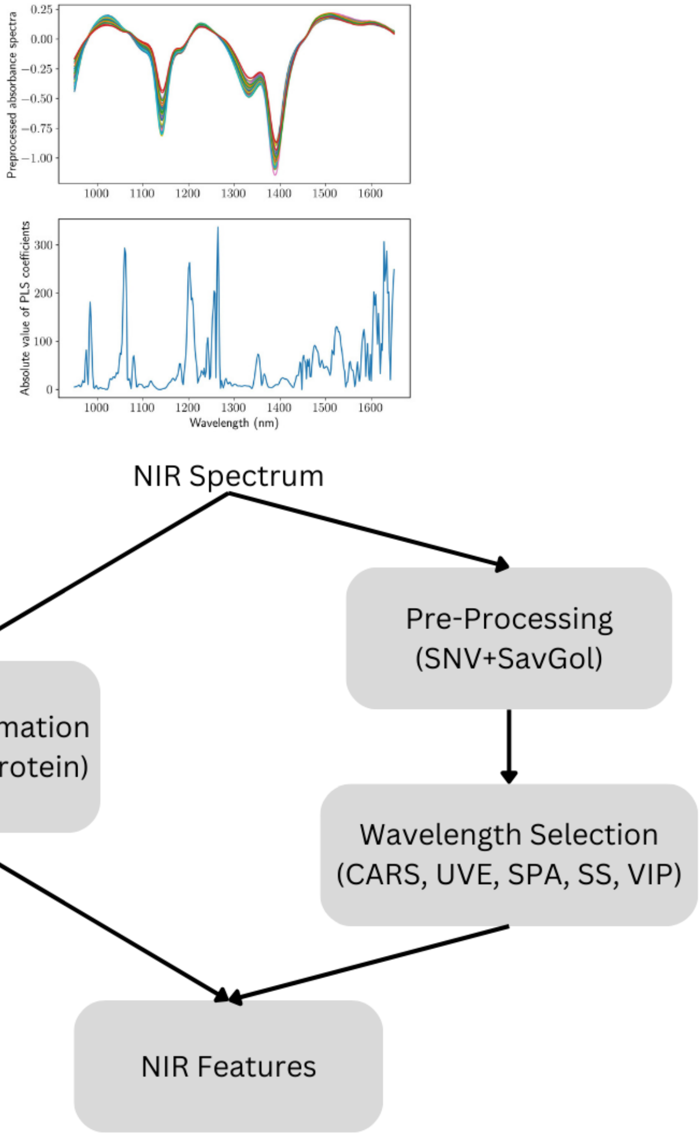


Figure 4.3: Schematic representation of NIR-based feature selection. The NIR spectrum undergoes preprocessing (SNV and Savitzky-Golay filtering) before wavelength selection using multiple techniques (CARS, UVE, SPA, SS, VIP). The selected spectral features contribute to CSPS estimation.

- Random Forest importance ranking, where each feature's contribution to prediction accuracy is estimated based on its frequency and effectiveness in tree-based splits.
- Variance Inflation Factor (VIF), to identify and remove collinear features that could introduce redundancy and noise in the model. A threshold of VIF ¿ 10 was used to

eliminate highly correlated predictors. If no features were returned from this process, VIF is skipped.

This combined approach enabled the identification of the most informative predictors, balancing both linear and nonlinear associations with CSPS.

Feature selection was evaluated at different scales, retaining the top 5, 10, and 15 features. To address potential collinearity, Principal Component Analysis (PCA) was optionally applied to these reduced feature sets, transforming them into 2, 3, or 5 orthogonal components before model training.

Model Selection and Calibration

To optimize the performance of our predictive models, we conducted an extensive hyperparameter tuning process using grid search cross-validation. The goal was to identify the best hyperparameters for each machine learning model—Ordinary Least Squares (OLS) [26], Random Forest (RF) [3], Support Vector Machine (SVM) [12], XGBoost (XGB) [5], and Gaussian Process Regression (GPR) [32]—based on their predictive accuracy and generalizability.

Each model was trained using three feature sets: RGB-Only, NIR-Only, and Fused Features (RGB + NIR), as described in Section 4.2.4. This approach allowed us to evaluate whether multimodal data fusion improved CSPS estimation compared to unimodal models.

Prior to training, we removed outliers that deviated more than two standard deviations from the mean for each of the laboratories; more information from the labs can be found at [6]. Furthermore, we removed samples that were outside the limits of agreements between the laboratories from a Bland-Altmann plot.

OLS regression is a fundamental statistical method used for modeling the relationship between independent variables and the dependent variable. Unlike other models, OLS does not require hyperparameter tuning but serves as a baseline to compare the performance of more complex models.

Random Forest models require tuning of parameters such as the number of trees, tree depth, and the number of features considered for each split. The results of our grid search cross-validation for RF are presented in Table 4.2.

Table 4.2: Random Forest grid search parameters used for hyperparameter tuning. The table outlines the tested values for key parameters, including the number of estimators, maximum depth, minimum samples split, minimum samples leaf, and maximum features. These combinations were explored to optimize the model's performance.

	Values		
Parameter	1	2	3
Number of Estimators	100	200	-
Maximum Depth	None	10	20
Minimum Samples Split	2	5	10
Minimum Samples Leaf	1	2	4
Maximum Features	sqrt	log2	None

For SVM models, critical hyperparameters include the kernel type, regularization parameter, and kernel coefficient (`gamma`) when using RBF kernels. The grid search results for SVM optimization are summarized in Table 4.3.

Table 4.3: Support Vector Machine (SVM) grid search parameters used for hyperparameter tuning. The table lists the values tested for key parameters, including the regularization parameter, kernel type, gamma value, and epsilon. These parameter combinations were evaluated to achieve optimal SVM model performance.

	Values		
Parameter	1	2	3
Regularization Parameter	0.1	1	10
Kernel	Linear	RBF	-
Gamma	Scale	Auto	-
Epsilon	0.1	0.2	0.5

XGBoost, a gradient boosting framework, was tuned by adjusting parameters such as the learning rate, maximum tree depth, and the number of boosting rounds. The results from the hyperparameter search for XGBoost are provided in Table 4.4.

Gaussian Process Regression (GPR) was also evaluated due to its probabilistic formulation, which naturally models prediction uncertainty—a desirable property for biological systems with noisy measurements. GPR models were optimized by testing combinations of

Table 4.4: XGBoost grid search parameters used for hyperparameter tuning. The table details the range of values tested for key parameters, including the number of estimators, learning rate, maximum depth, subsample ratio, feature selection percentage, Lasso and Ridge regularization strengths, and minimum loss reduction. These combinations were evaluated to optimize the model's performance.

	Values		
Parameter	1	2	3
Number of Estimators	100	200	-
Learning Rate	0.1	0.01	-
Maximum Depth	None	10	20
Subsample	0.6	0.8	1.0
Feature Selection Percentage	0.6	0.8	1.0
Lasso Regularization	0	0.1	1.0
Ridge Regularization	1	5	10
Minimum Loss Reduction	0.0	0.1	1.0

kernel functions, including Dot Product, Radial Basis Function (RBF), and White Noise kernels.

These tuning results guided the selection of the best model configurations, which were subsequently used for final model evaluation and comparison.

Model Evaluation Strategy

To evaluate model performance, we implemented a nested 5-fold cross-validation (CV) framework, where the inner loop performed hyperparameter tuning while the outer loop assessed the model's generalizability. This approach mitigates data leakage and ensures an unbiased estimate of predictive performance. To evaluate we concatenated the test results of each fold and evaluated the R^2 to measure the variance explainability of the model and with a maximum value of 1.0, indicating perfect explainability. We also measured the correlation of the predicted CSPS with the laboratory estimate.

Data Fusion Strategy

To improve CSPS prediction, we explored two data fusion approaches that integrate features from RGB image analysis and near-infrared (NIR) spectroscopy. The motivation behind data fusion is that image-based features capture physical kernel size distributions, while NIR spectroscopy provides chemical composition data (starch, fiber, protein), allowing for a more comprehensive assessment of silage processing.

We implemented two fusion strategies to integrate NIR and image-derived features for CSPS prediction: feature-level fusion (early fusion) and decision-level fusion (late fusion).

- RGB-Only Features
- NIR-Only Features
- Early Fusion (RGB + NIR Features)
- Late Fusion (RGB + NIR Predictions)

The RGB-Only Features approach relies on image-based particle size metrics extracted from RGB images, including Maximum Inscribed Circle Diameter (MICD), area, shape descriptors, and CSPS values estimated using a previously established equation. The NIR-Only Features strategy utilizes wavelength-selected spectral data obtained through various feature selection techniques such as PLSR-VIP, CARS, SPA, UVE, and PCA. For the Early Fusion method, selected RGB and NIR features were concatenated into a unified feature vector, which was then used to train a single regression model for CSPS prediction. In contrast, the Late Fusion approach involves training separate regression models on RGB and NIR features independently; their respective predictions are then combined through a secondary regression model to generate the final CSPS estimation.

Early Fusion In early fusion, feature selection was performed jointly on the combined dataset of RGB and NIR features, and the selected features were used to train a single regression model. This approach assumes that both data modalities are complementary and that their integration at the feature level enhances predictive accuracy.

Late Fusion In late fusion, separate models were trained using RGB and NIR features independently. The predictions from these models were then used as inputs to a second-stage regression model that produced the final CSPS prediction. This approach allows

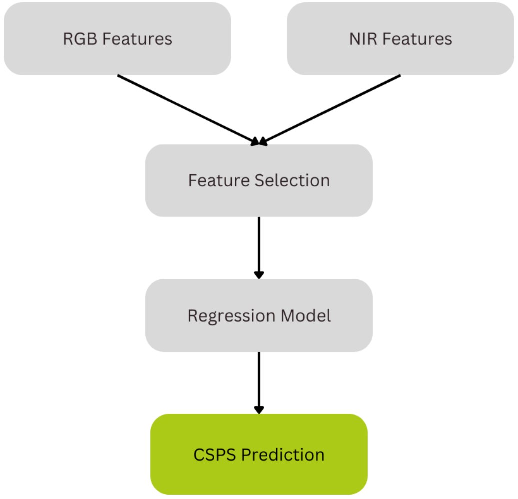


Figure 4.4: Early fusion strategy: RGB and NIR features are combined before feature selection and fed into a single regression model for CSPS prediction.

each modality to contribute separately, capturing distinct patterns before merging their outputs.

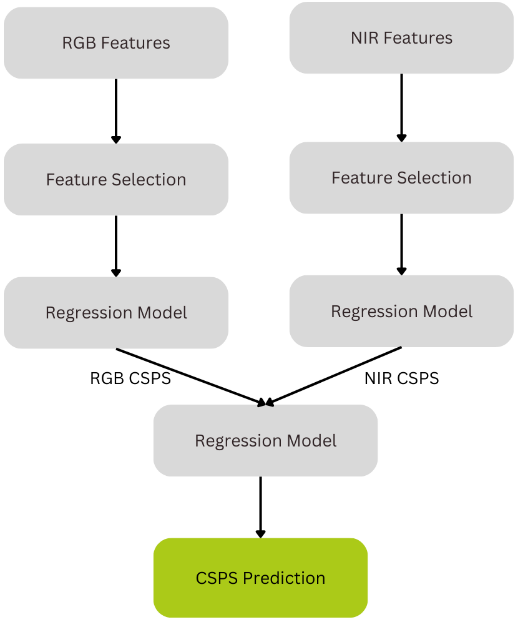


Figure 4.5: Late fusion strategy: RGB and NIR features are used independently to train separate models, and their predictions are later combined to estimate CSPS.

To assess the effectiveness of these fusion strategies, models were trained using RGB-only features, NIR-only features, early fusion, and late fusion. Performance was evaluated based on cross validated RMSE - $RMSE_{CV}$ - and R^2. $RMSE_{CV}$ quantifies the average prediction error across cross-validation folds, offering a robust estimate of how well the model generalizes to unseen data. A lower $RMSE_{CV}$ indicates greater accuracy and reliability in CSPS prediction. Meanwhile, R^2 measures the proportion of variance in the reference CSPS values explained by the model, with values closer to 1 denoting better explanatory power. Together, these metrics enable a comprehensive comparison of unimodal and multimodal strategies, helping to identify which fusion method yields the most accurate and informative CSPS estimates.

4.2.5 Reference Laboratory and Optical CSPS Analysis

Details regarding laboratory CSPS measurement protocols, reference values, and inter-laboratory agreement analyses can be found in [6]. The removal of the outliers followed two distinct criteria: First, three plots were excluded due to excessive disagreement between the two reference laboratories, where the difference in CSPS exceeded 13.04, corresponding to the 95% limit of agreement determined by Bland-Altman analysis. This level of disagreement suggests potential measurement inconsistencies or sampling errors. The second criterion removed three additional plots with CSPS values that fell outside two standard deviations of the mean, ensuring that extreme observations with limited biological or practical relevance did not distort model training.

To better understand the reliability of CSPS measurements and their impact on model performance, we evaluated the signal-to-noise ratio using a comparison between laboratories. For this analysis, three paired plots with identical machine settings were measured by both laboratories to estimate the standard error of the laboratory (SEL). The SEL was be compared to the observed sample standard deviation to quantify measurement uncertainty and assess its influence on CSPS predictions.

In parallel, we evaluated the signal-to-noise ratio for particle sizing metrics, including the geometric mean particle size of the grain (GMPS) and the fraction of the grain (FoG), both of which are known to correlate strongly with the CSPS. For this analysis, duplicate measurements were performed on 59 plots, enabling SEL estimation for both GMPS and FoG. A lower ratio of SEL to sample standard deviation would indicate reduced measurement noise and, consequently, more reliable inputs for model development. Given their biological relevance and correlation with CSPS, improved signal quality in FoG and GMPS may contribute to models with higher predictive power and lower unexplained variance. Moreover, morphological features extracted from RGB data might have better correlation to FoG, meanwhile, chemical signatures captured by NIR might have better correlation with CSPS.

Finally, to evaluate the predictive accuracy of the developed models, we calculated the Pearson's correlation coefficient between the predicted Optical CSPS scores and the laboratory-measured CSPS values. This correlation assess the strength of association between model estimates and reference standards.

4.3 Results and Discussion

This section presents the results obtained from the image analysis, feature selection, and predictive modeling processes. It begins with an overview of the CSPS distribution across all plots, followed by an analysis of the impact of feature selection strategies on model performance. Particular attention is given to the role of machine settings, specifically Roll Gap, and the ability of models to generalize when these settings are unavailable. Finally, the performance of the best-performing models is summarized, highlighting the influence of image-derived metrics, spectral features, and machine parameters on predictive accuracy.

4.3.1 Dataset

Figure 4.6 shows the images with most values corn kernels in the plots with lowest and highest CSPS. Figure 4.7 illustrates the distribution of CSPS values across all plots after the removal of six outliers.

The final dataset exhibits a relatively symmetric distribution, with a mean CSPS of 75.24 and a median of 76.53, indicating minimal skewness. Approximately 68% of plots

(a) Sample from the plot with lowest CSPS (59.9%)

(b) Sample from the plot with highest CSPS (89.3%)

Figure 4.6: Representative images from plots with contrasting corn silage processing scores (CSPS). Panel (a) corresponds to the lowest CSPS sample, characterized by larger and more intact kernel fragments. Panel (b) shows the highest CSPS sample, exhibiting finer particle fragmentation and fewer whole kernels. These images highlight the morphological differences associated with processing quality.

fall within one standard deviation of the mean (67.06 to 83.42), and 95% fall within two standard deviations (58.88 to 91.60). This dispersion reflects the expected variability introduced by differences in harvest date, forage harvester settings, and field conditions. The 25th percentile (70.21) and 75th percentile (81.55) define the interquartile range, emphasizing the diversity of CSPS outcomes observed in the field.

Following particle segmentation and annotation, the quantiles of particle size across all samples were: Q25 = 1.27 mm, Q50 = 1.90 mm, Q75 = 2.90 mm, Q90 = 3.96 mm, Q95 = 4.75 mm, Q97 = 5.24 mm, and Q99 = 5.88 mm. Notably, the Q95 value aligns with the industry-standard sieve threshold used in CSPS lab protocols, reinforcing its predictive value for regression models. Accordingly, particle counts and cumulative area fractions above the Q95 threshold were prioritized as features in the image-based CSPS estimation.

To assess the reliability of the laboratory reference CSPS, inter-laboratory comparison

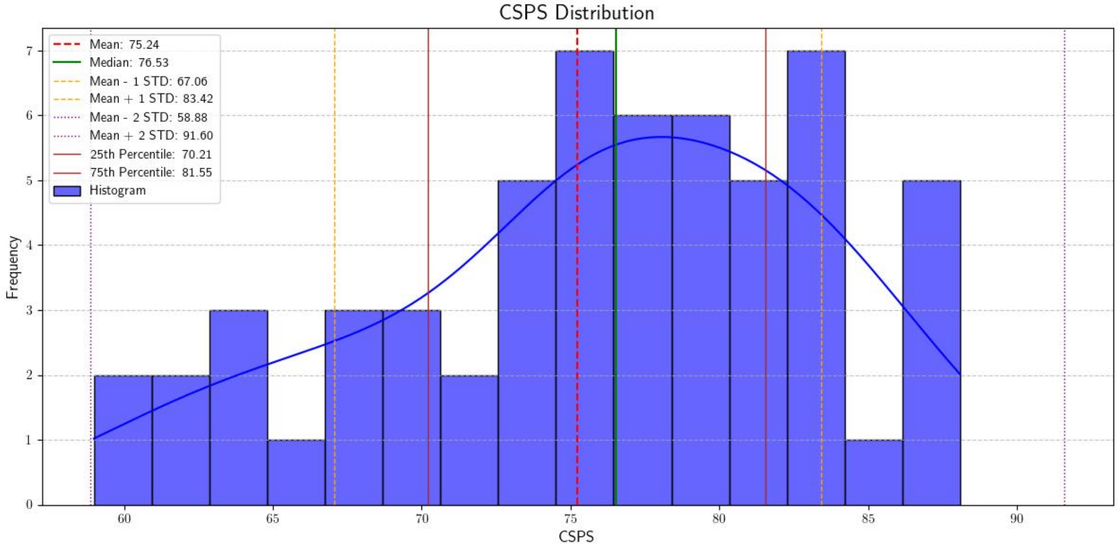


Figure 4.7: Distribution of Corn Silage Processing Scores (CSPS) after outlier removal. The red dashed line represents the mean (75.24), the green solid line denotes the median (76.53), yellow dashed lines indicate one standard deviation (67.06, 83.42), and purple dotted lines indicate two standard deviations (58.88, 91.60). The blue solid curve represents a smoothed density estimate. Vertical markers for the 25th percentile (70.21) and 75th percentile (81.55) further characterize the distribution.

was conducted using 30 plots measured by two independent laboratories. The Standard Error of Laboratory (SEL) for Lab 1 was 6.7%, and for Lab 2 it was 6.9%. Given the overall standard deviation of CSPS was 6.76%, the noise-to-signal ratio was high, suggesting that regression models trained on this reference may face limitations in capturing true processing performance. The Concordance Correlation Coefficient (CCC) between labs was 0.689, reflecting moderate agreement. Bland-Altman analysis (Figure 4.8) indicates better inter-lab agreement for highly processed samples and smaller roll gap clearances, where particle fragmentation is more uniform and sampling error is reduced.

In contrast, the Fraction of Grain (FoG) greater than 4.75 mm, derived from image analysis, exhibited a standard deviation of 8.51% and an SEL of 2.52% based on duplicate measurements from 59 plots. This suggests a more favorable signal-to-noise ratio compared to CSPS. Given FoG's high correlation with CSPS, its stability supports its utility as a proxy in optical assessments.

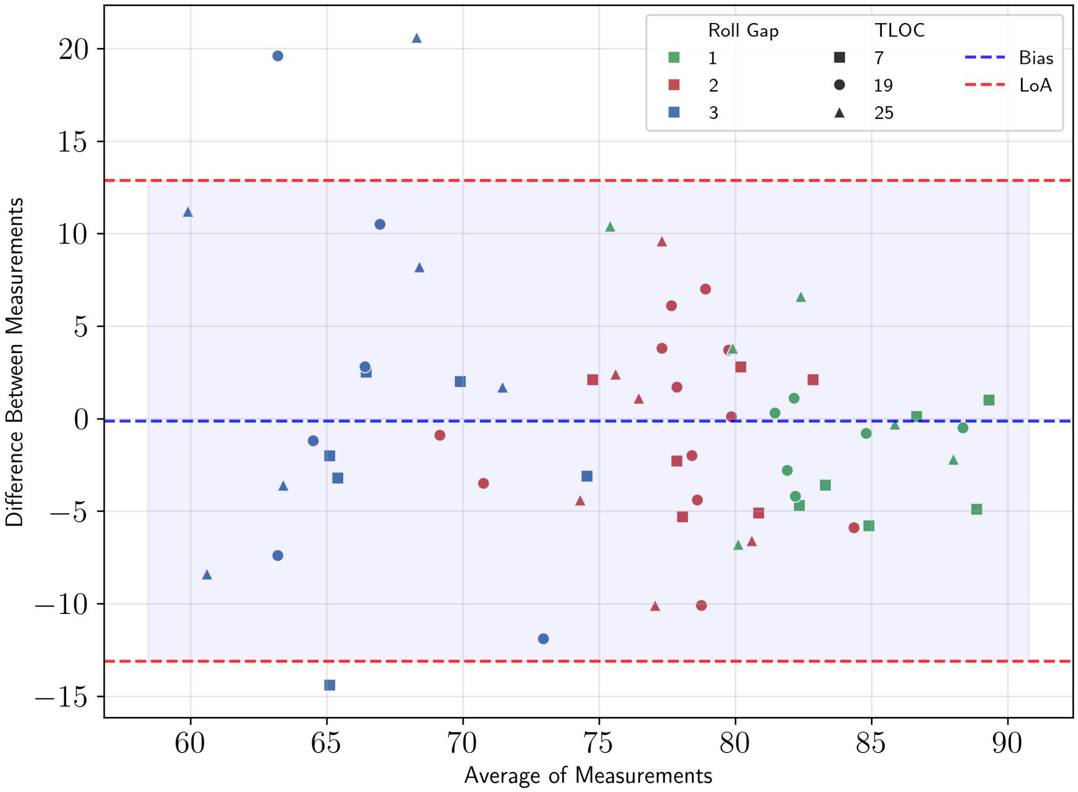


Figure 4.8: Bland-Altman plot of CSPS measurements from two independent laboratories. Colors show Roll Gap levels (green = 1, red = 2, blue = 3); shapes indicate TLOC (square = 7 mm, circle = 19 mm, triangle = 25 mm). Blue dashed line indicates mean bias, red dashed lines indicate limits of agreement (LoA). Agreement improves in highly processed samples, suggesting more homogeneous fragmentation reduces sampling error.

4.3.2 Optical CSPS Correlation

This section evaluates the potential of using image-derived features as proxies for laboratory CSPS by analyzing their correlation with reference scores. The focus is on particle size metrics that reflect kernel fragmentation, particularly those capturing area, volume, and shape characteristics of large particles.

Figure 4.9 demonstrates the strong negative relationship between the fraction of grain particles larger than 4.75 mm and the laboratory CSPS. The regression line indicates that increases in large, unprocessed particles are consistently associated with lower CSPS values. The fitted linear model yielded an R^2 of 0.853, Pearson's r of -0.923, and a highly significant p-value ($\ll 0.05$), confirming the robustness of the association.

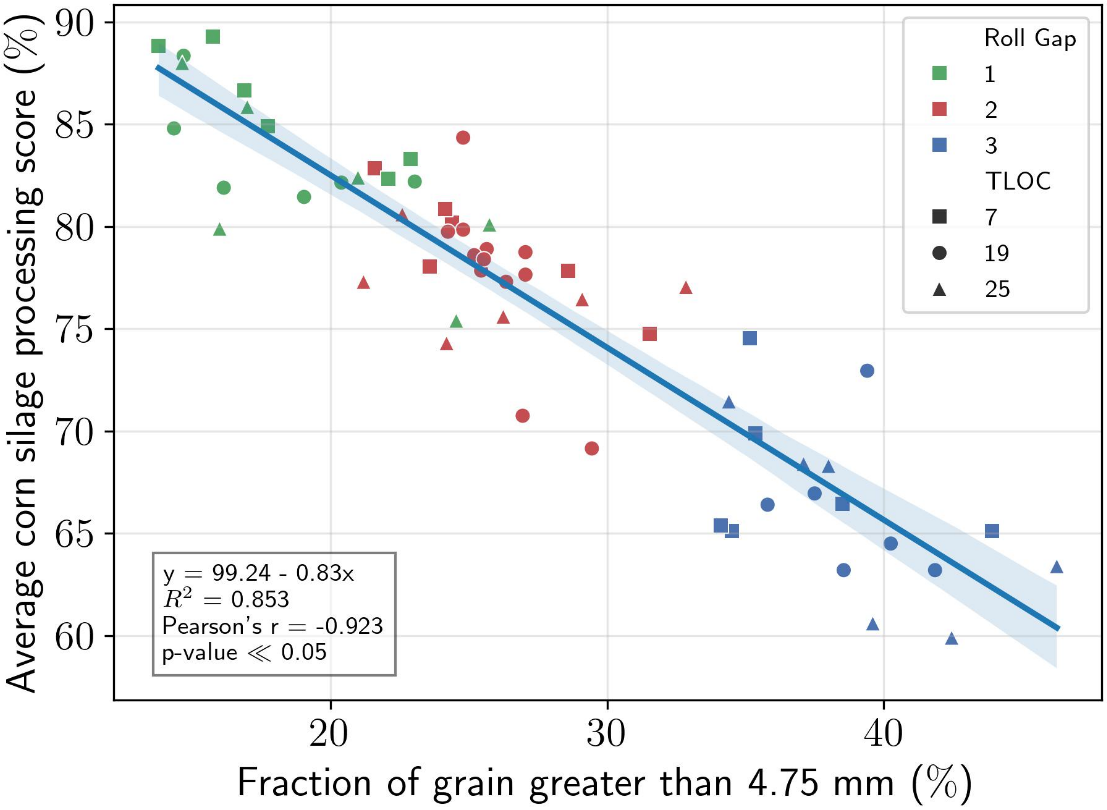


Figure 4.9: Linear regression between the laboratory CSPS and the fraction of grain particles larger than 4.75 mm. Each point's color represents the Roll Gap (green = 1.0, red = 2.0, blue = 3.0), and shape represents the Theoretical Length of Cut (TLOC: square = 7 mm, circle = 19 mm, triangle = 25 mm). The solid blue line is the fitted regression line, and the shaded blue area indicates the 95% confidence interval. A strong inverse relationship is observed, with $R^2 = 0.853$, $r(p \ll 0.05) = -0.923$.

Drewry et al. (2019) [7] utilized the maximum inscribed circle diameter (MICD) as a key particle size metric, whereas Rasmussen et al. (2019) [24] employed the shortest axis length to characterize kernel dimensions to apply to calculate the CSPS through equation 4.3 [10]. Drewry's study was conducted in a controlled environment using hydrodynamically separated corn kernels on a black background, while Rasmussen analyzed static harvested silage, making the latter's imaging conditions more comparable to our dataset. However, Rasmussen's study lacks a direct comparison between their optical CSPS estimates and Ro-Tap/laboratory reference values, whereas Drewry provides such an evaluation. This gap highlights the need to investigate which kernel size metric for particles under 4.75 mm best correlates with laboratory-measured CSPS in corn silage images.

$$CSPS = 100 - \frac{\%Starch_{>4.75}}{\%Starch_{total}} \tag{4.3}$$

Additionally, to estimate starch content, Drewry et al. (2019) [7] used an area-to-weight conversion, but alternative metrics, such as volume-based estimations, could provide a more accurate assessment. Thus, to explore optical CSPS estimation, we systematically evaluated multiple metric combinations.

To further assess the predictive value of individual optical features, we computed Pearson's correlation between CSPS and each image-derived metric. Table 4.5 summarizes the five most strongly correlated features with laboratory CSPS.

Table 4.5: Top five image-derived metrics most correlated with laboratory CSPS

Optical Metric	**Pearson's r**
CSPS Area $<$ Q97	0.588
CSPS Area $<$ Q90	0.542
CSPS MICD $<$ Q95	0.540
CSPS Area $<$ Q95	0.537
CSPS Volume $<$ Q95	0.524

Among these, cumulative particle area below the 97th percentile (5.24 mm) showed the strongest correlation, reinforcing its relevance for optical CSPS estimation. Metrics using thresholds near the Q95 sieve cut-off (4.75 mm), such as CSPS Area ¡ Q95 and MICD ¡ Q95, also exhibited strong performance, consistent with expectations from standard CSPS lab definitions. It is also relevant to mention that all of the metrics in Table 4.5 are highly correlated ($r > 0.95$)

Figures 4.10 and 4.11 show the predicted versus observed CSPS and the residuals of the best univariate model based on CSPS Area $<$ Q97. Although this single feature captures relevant size information, a moderate residual spread remains, highlighting the need for multivariate models to fully account for variance in processing scores.

These findings are consistent with those reported by Drewry et al. [7], who demonstrated

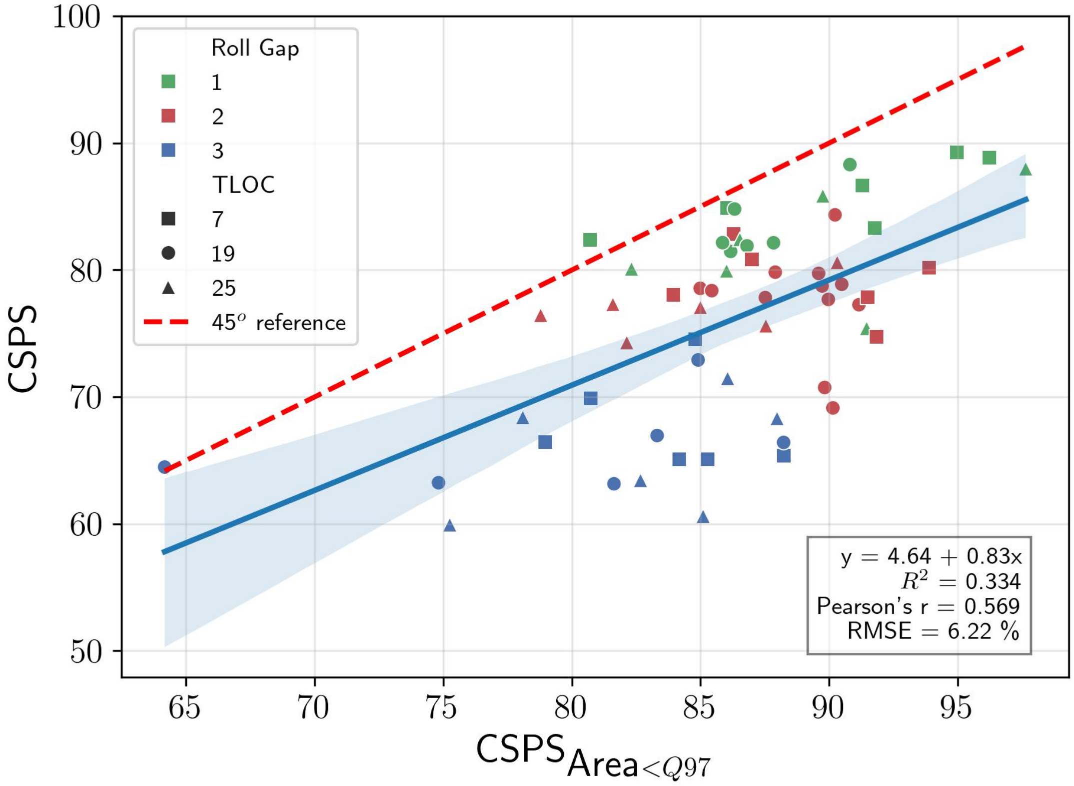


Figure 4.10: Observed CSPS versus optical estimates based on CSPS Area < Q97. Colors show Roll Gap levels (green = 1, red = 2, blue = 3); shapes indicate TLOC (square = 7 mm, circle = 19 mm, triangle = 25 mm). The blue line shows the regression fit with a shaded 95% confidence band. The red dashed line denotes the 1:1 ideal prediction.

that image-derived kernel size metrics such as cumulative undersize percent below 4.75 mm were highly correlated with laboratory-based CSPS values. In their study, a Pearson correlation of $r = 0.8$ was observed across 28 corn silage samples using cross-sectional area-based segmentation. Our results corroborate the utility of area-based thresholds and extend their applicability to a larger dataset with high-speed imagery. The inclusion of multivariate optical models further strengthens the case for using vision-based proxies as scalable alternatives to labor-intensive laboratory measurements.

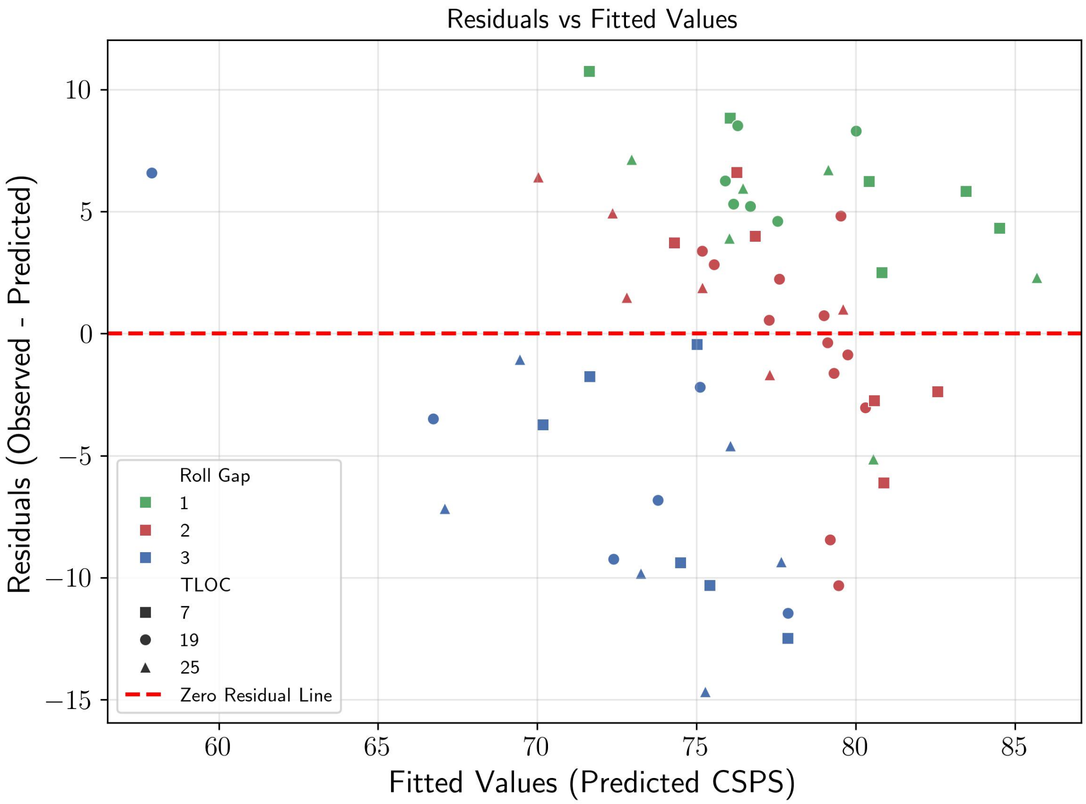


Figure 4.11: Residual plot of the regression using CSPS Area < Q97. Colors show Roll Gap levels (green = 1, red = 2, blue = 3); shapes indicate TLOC (square = 7 mm, circle = 19 mm, triangle = 25 mm). The blue line shows the regression fit with a shaded 95% confidence band. The red dashed line denotes the 1:1 ideal prediction. The dashed red line is the 45° reference line representing perfect prediction.

4.3.3 Machine Learning CSPS

RGB Features Regression

The best-performing RGB-based regression model predicted the CSPS using Fragment of Grain (FoG) as a proxy ($R^2 = 0.727$, Pearson's r = 0.856, $RMSE_{CV} = 2.61\%$) compared to direct CSPS estimation ($R^2 = 0.290$, Pearson's r = 0.551, $RMSE_{CV} = 6.21\%$). This model leveraged features selected via Pearson correlation, retaining the top 15 features. Dimensionality was further reduced using Principal Component Analysis (PCA), retaining 3 components. The model itself was an XGBoost regressor, tuned with the following key hyperparameters: 50 estimators, a learning rate of 0.1, maximum tree depth of 3, subsample and column sample ratios both set to 0.8, and regularization parameters $\alpha = 0$ and $\lambda = 1$.

This RGB-based approach showed competitive performance by effectively capturing visual cues related to fragment of grain above 4.75 mm, which strongly correlates with CSPS. However, as shown in Figure 4.12, direct NIR-based estimation ultimately yielded better accuracy due to its chemical specificity, as discussed in the following section.

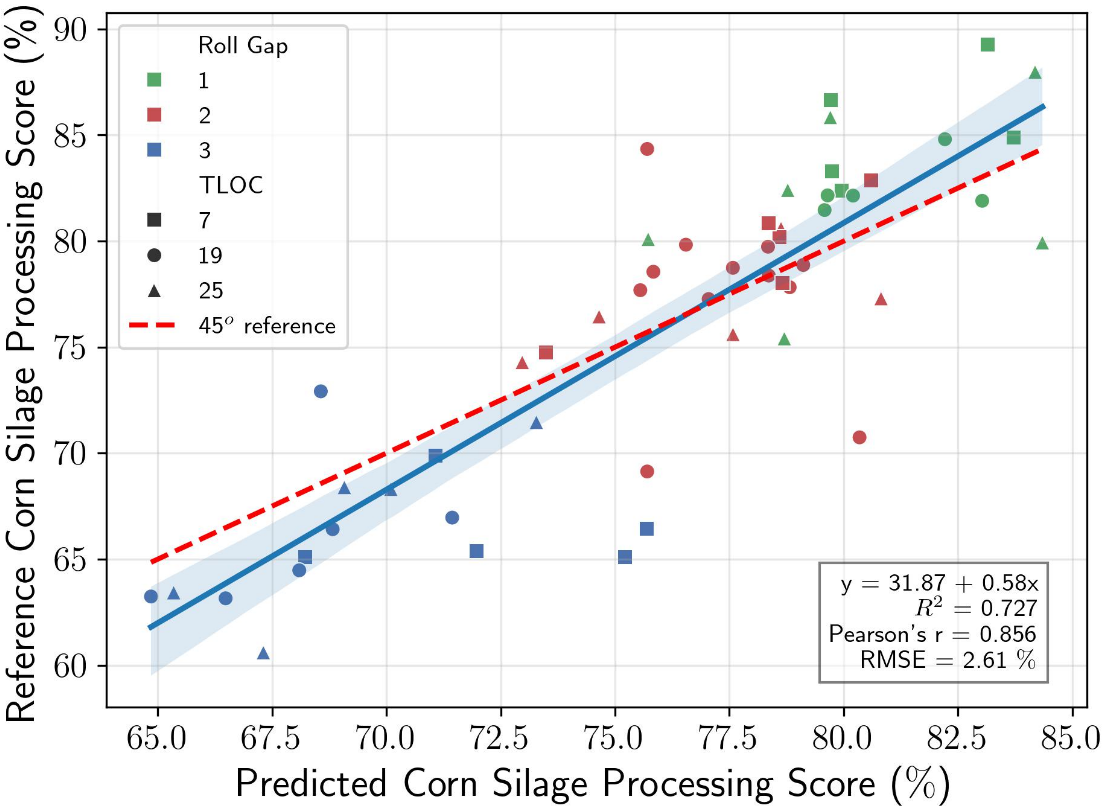


Figure 4.12: RGB-derived features were first used to predict Fragment of Grain (FoG), which was then mapped to CSPS using a calibrated regression. Colors show Roll Gap levels (green = 1, red = 2, blue = 3); shapes indicate TLOC (square = 7 mm, circle = 19 mm, triangle = 25 mm). The blue line shows the regression fit with a shaded 95% confidence band. The red dashed line denotes the 1:1 ideal prediction. Fit statistics include the regression equation, R^2, Pearson's r, and $RMSE_{CV}$.

The top five most influential features for the RGB-based model were: (1) the mean particle area, (2) the count of particles with area greater than the 97th quantile (corresponding to approximately 5.24 mm), (3) the standard deviation of the major axis length, (4) the computed CSPS using a volume-based starch estimation, and (5) the starch volume from particles with major axis lengths below the 97th quantile threshold. As anticipated, metrics capturing particle size distribution—particularly mean and standard deviation—played a dominant role in prediction performance. Consistent with findings from Rocha et al. [25], the number of large particles had a significant impact on CSPS. However, we found a 5.24 mm threshold to yield stronger predictive relevance than the conventional 4.75 mm cutoff. This suggests that under more complex image-derived quantification scenarios, using FoG as proxy might introduce a systematic bias, potentially necessitating threshold reevaluation.

NIR Features Regression

The best-performing NIR-based regression model directly estimated the Corn Silage Processing Score (CSPS) ($R^2 = 0.911$, Pearson's r = 0.902, $RMSE_{CV} = 2.69\%$) compared with relying on proxy indicators ($R^2 = 0.001$, Pearson's r = -0.154, $RMSE_{CV} = 7.40\%$). Recursive Feature Elimination (RFE) was used to select the 15 most informative NIR features, and no VIF-based collinearity filtering was applied. To reduce dimensionality, Principal Component Analysis (PCA) was performed, retaining 5 principal components. The regression model used was a Random Forest Regressor with 50 estimators, unrestricted tree depth, a minimum of 2 samples required to split an internal node, and a minimum of 1 sample per leaf.

As shown in Figure 4.13, this approach achieved robust predictive accuracy, capitalizing on the chemical signatures captured by the NIR spectrum to estimate CSPS more precisely than image-derived proxies. The absence of a proxy (FoG) features makes this pipeline more directly interpretable for applications targeting physical quality measures of forage processing.

The NIR-based model derived its predictions from a diverse combination of spectral features selected using Recursive Feature Elimination (RFE). The top contributors, as identified by SHAP analysis, included wavelengths selected by methods such as VIP

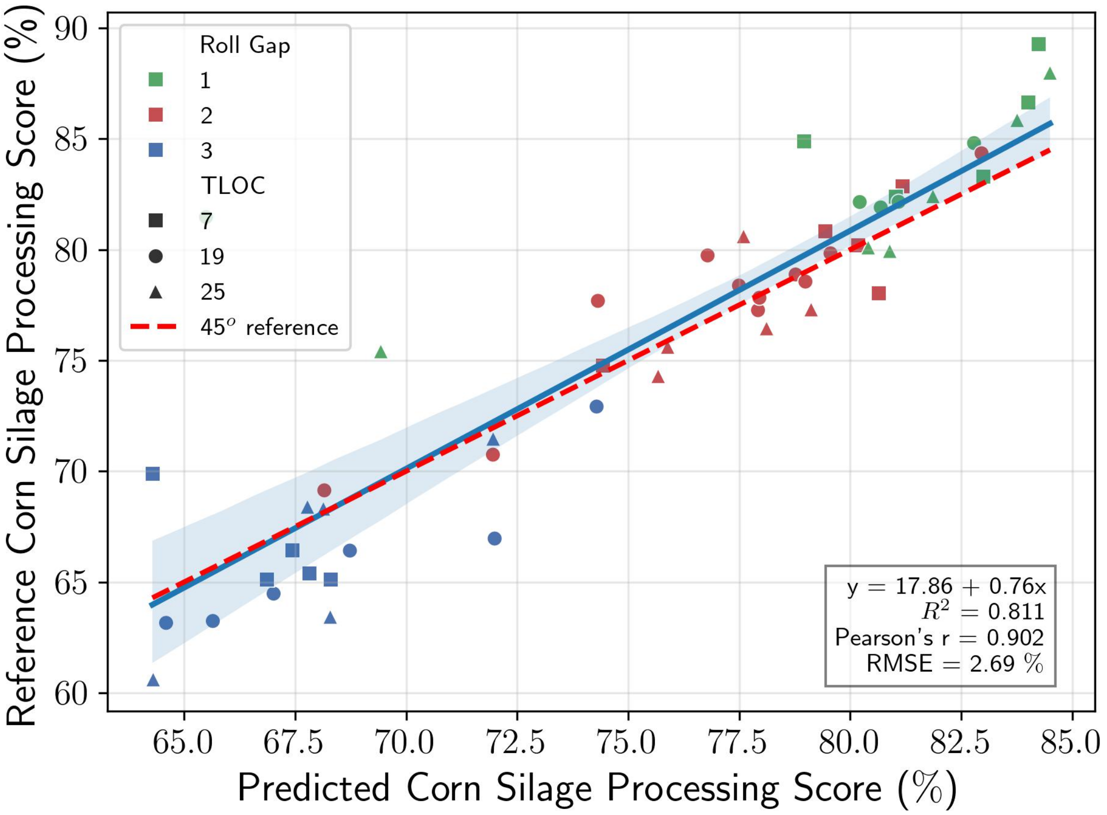


Figure 4.13: CSPS prediction using direct regression from NIR spectral estimations. Colors show Roll Gap levels (green = 1, red = 2, blue = 3); shapes indicate TLOC (square = 7 mm, circle = 19 mm, triangle = 25 mm). The blue line shows the regression fit with a shaded 95% confidence band. The red dashed line denotes the 1:1 ideal prediction. Fit metrics including R^2, Pearson r, and $RMSE_{CV}$ are shown in the annotation box.

(Variable Importance in Projection), CARS (Competitive Adaptive Reweighted Sampling), SPA (Successive Projection Algorithm), and SS (Stepwise Selection). In particular, features such as VIP1202.0, CARS1182.0, and CARS1060.0 had the strongest impact on the model output. This consistent trend suggests that absorbance at these wavelengths is positively correlated with well-processed silage.

Other influential wavelengths included those near 1062 nm, 1156 nm, and 1370 nm, regions often associated with chemical bonds related to fiber, starch, and protein absorption in NIR spectroscopy reflectance. The dominance of VIP-selected features across several bands underscores the importance of chemically meaningful regions in modeling CSPS.

Early Fusion Regression

For the early fusion strategy, all RGB and NIR features were concatenated into a unified feature set and used to directly predict the laboratory-measured CSPS ($R^2 = 0.568$, Pearson's r = 0.759, $RMSE_{CV} = 5.90\%$), bypassing the use of any intermediate proxy ($R^2 = 0.549$, Pearson's r = 0.745, $RMSE_{CV} = 5.32\%$). The top 15 features were selected using Mutual Information-based ranking, and no collinearity filtering (VIF) or PCA dimensionality reduction was applied. The best performance was achieved using a Gaussian Process Regressor (GPR) with a Dot Product kernel and a White noise kernel component, specifically defined as $\text{DotProduct} + \text{WhiteKernel}(noise\_level = 5.0,\ noise\_level\_bounds = (1.0,\ 25.0))$.

Despite leveraging both imaging and spectroscopic data, the early fusion model did not outperform the best individual RGB and NIR models. As shown in Figure 4.14, its predictive performance was lower in both R^2 and $RMSE_{CV}$ metrics. This suggests that combining heterogeneous feature sets may introduce redundancy or noise, and highlights the importance of tailored modeling strategies or late fusion methods to fully exploit the complementary nature of RGB and NIR data.

The Early Fusion model combined both RGB-derived morphological features and NIR spectral data to predict the Corn Silage Processing Score (CSPS). While the integration of these modalities aimed to leverage complementary information—structural indicators from images and compositional cues from NIR—the fused model did not outperform the best unimodal approaches. This suggests that simply concatenating features from different sensors may not be sufficient to boost predictive performance and could even introduce redundancy or conflict when not appropriately calibrated.

Among the most influential features in the early fusion model were several NIR wavelengths associated with fiber and starch absorption, particularly in the 980–1344 nm range [1, 2, 29]. Moreover, particle metrics, such as the proportion of areas or minor axes below or above high quantiles (e.g., 97th), also had noticeable effects on predictions. This is consistent with established domain knowledge, wherein the physical mechanisms of particle

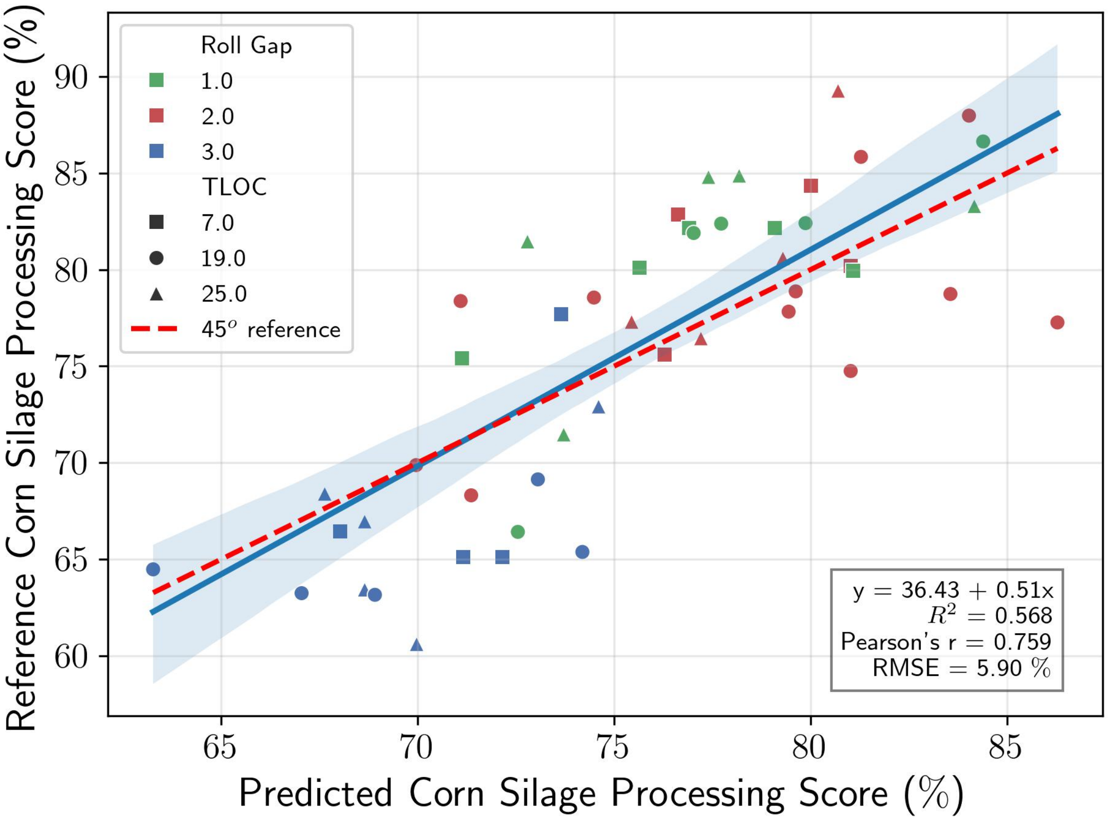


Figure 4.14: Predicted vs. reference CSPS using early fusion of RGB and NIR features. Colors show Roll Gap levels (green = 1, red = 2, blue = 3); shapes indicate TLOC (square = 7 mm, circle = 19 mm, triangle = 25 mm). The blue line shows the regression fit with a shaded 95% confidence band. The red dashed line denotes the 1:1 ideal prediction. Regression equation, R^2, Pearson's r, and $RMSE_{CV}$ are displayed.

fragmentation and kernel breakage are directly correlated with CSPS values.

The moderate performance of the fusion model relative to unimodal models highlights the importance of thoughtful data integration. Notably, recent studies have identified specific challenges associated with early fusion strategies, especially when dealing with small datasets. For instance, early fusion can lead to overfitting due to the increased complexity from combining heterogeneous data sources, and the potential for one modality to dominate the learning process, thereby diminishing the benefits of multimodal integration [31, 34]. These findings underscore the need for more sophisticated fusion techniques or alternative strategies, such as late fusion, which can better manage the unique characteristics and contributions of each modality.

Late Fusion Regression

Late fusion was evaluated by combining predictions from the best-performing RGB and NIR models. In this study, the RGB model was trained to predict a proxy variable—the Fragment of Grain (FoG)—while the NIR model directly predicted the Corn Silage Processing Score (CSPS). The combined predictions were fused using several techniques, including weighted averaging, support vector regression (SVR), linear regression, and random forest regression.

To find the optimal weight α for the weighted fusion method, a linear search was performed between 0 and 1, with step of 0.01. As shown in Figure 4.15, the optimal fusion minimizing the $RMSE_{CV}$ occurred near $\alpha = 0.33$, giving more weight to the NIR model.

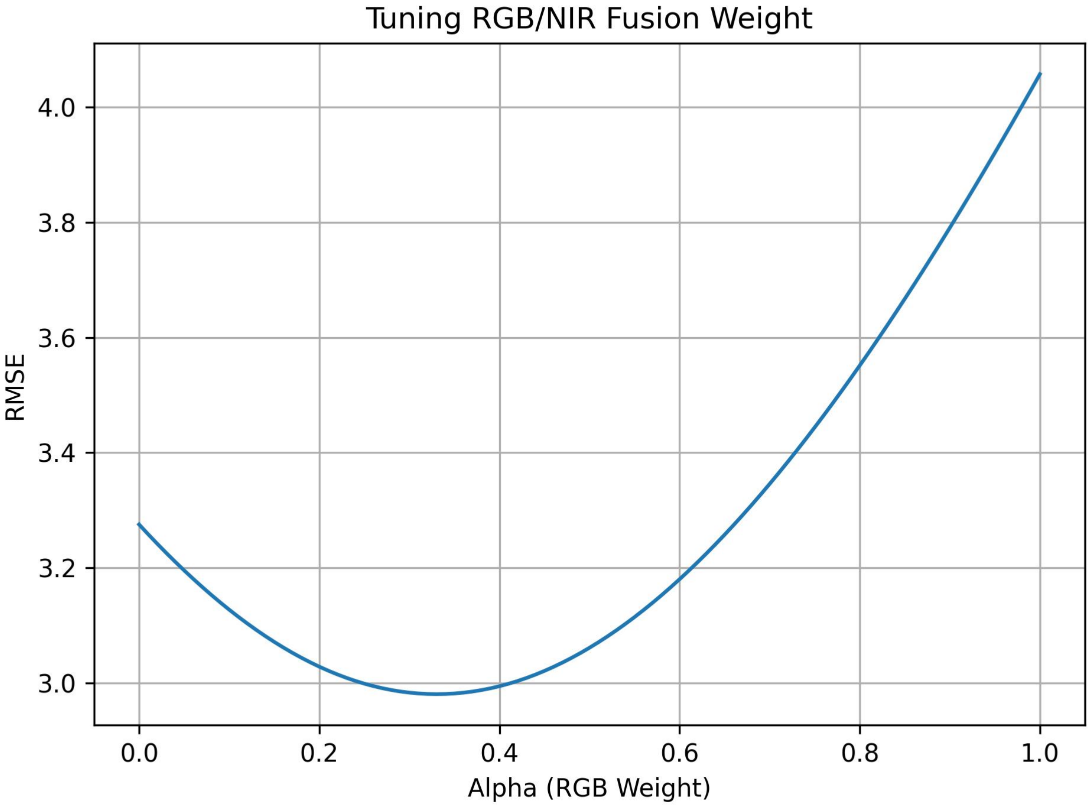


Figure 4.15: $RMSE_{CV}$ curve from tuning the late fusion weight α assigned to the RGB model output. The x-axis represents the RGB weight α (ranging from 0 to 1), while the NIR model weight is $1 - \alpha$. The y-axis shows the Root Mean Squared Error ($RMSE_{CV}$) between the fused prediction and the reference CSPS. The optimal weight minimizing $RMSE_{CV}$ lies around $\alpha \approx 0.3$, indicating a stronger contribution from the NIR model.

The results of the four late fusion strategies are illustrated in Figure 4.16. Overall, our findings show that combining NIR spectroscopy and RGB-derived image features through late fusion provides the most accurate prediction of laboratory CSPS, achieving a test R^2 of 0.949, $RMSE_{CV}$ of 1.68%, and Pearson's r of 0.976. This performance surpasses unimodal models (RGB-only: $R^2 = 0.713$, $RMSE_{CV} = 3.69\%$; NIR-only: $R^2 = 0.718$, $RMSE_{CV} = 3.77\%$) and early fusion ($R^2 = 0.568$, $RMSE_{CV} = 5.90\%$), highlighting the limitations of early feature concatenation and the advantage of preserving modality-specific structure.

Compared to previous studies, our model yields significantly improved predictive accuracy. Drewry et al. [7] achieved $r = 0.80$ when estimating CSPS from images of hand-separated kernels photographed under controlled conditions; in contrast, our best univariate proxy (FoG ¿ 4.75 mm) achieved $r = -0.92$ and $R^2 = 0.85$ under realistic field imaging conditions, and achieving $r = 0.856$ and $R^2 = 0727$ while estimating the CSPS. Rocha et al. [25], using RGB features and machine learning, reported R^2 values between 0.64 and 0.79 for predicting CSPS categories, though their models explored mostly whole kernel count to predict CSPS and lacked chemical information from NIR. Rasmussen and Moeslund [24] estimated kernel fragmentation in silage using object detection and reported classification F1-scores between 0.70 and 0.84; however, their pipeline did not estimate CSPS directly, and their dataset lacked ground truth from lab measurements.

Our study provides the first comprehensive integration of NIR and RGB data for CSPS regression validated against dual-lab measurements, showing higher precision and lower error than prior works. Additionally, our protocol explicitly quantifies signal-to-noise ratios for CSPS and FoG, and leverages robust modeling strategies to handle field variability and imaging noise.

4.4 Conclusion

This study presents a comprehensive framework for estimating the Corn Silage Processing Score (CSPS) using image-derived morphological features, near-infrared (NIR) spectral data,

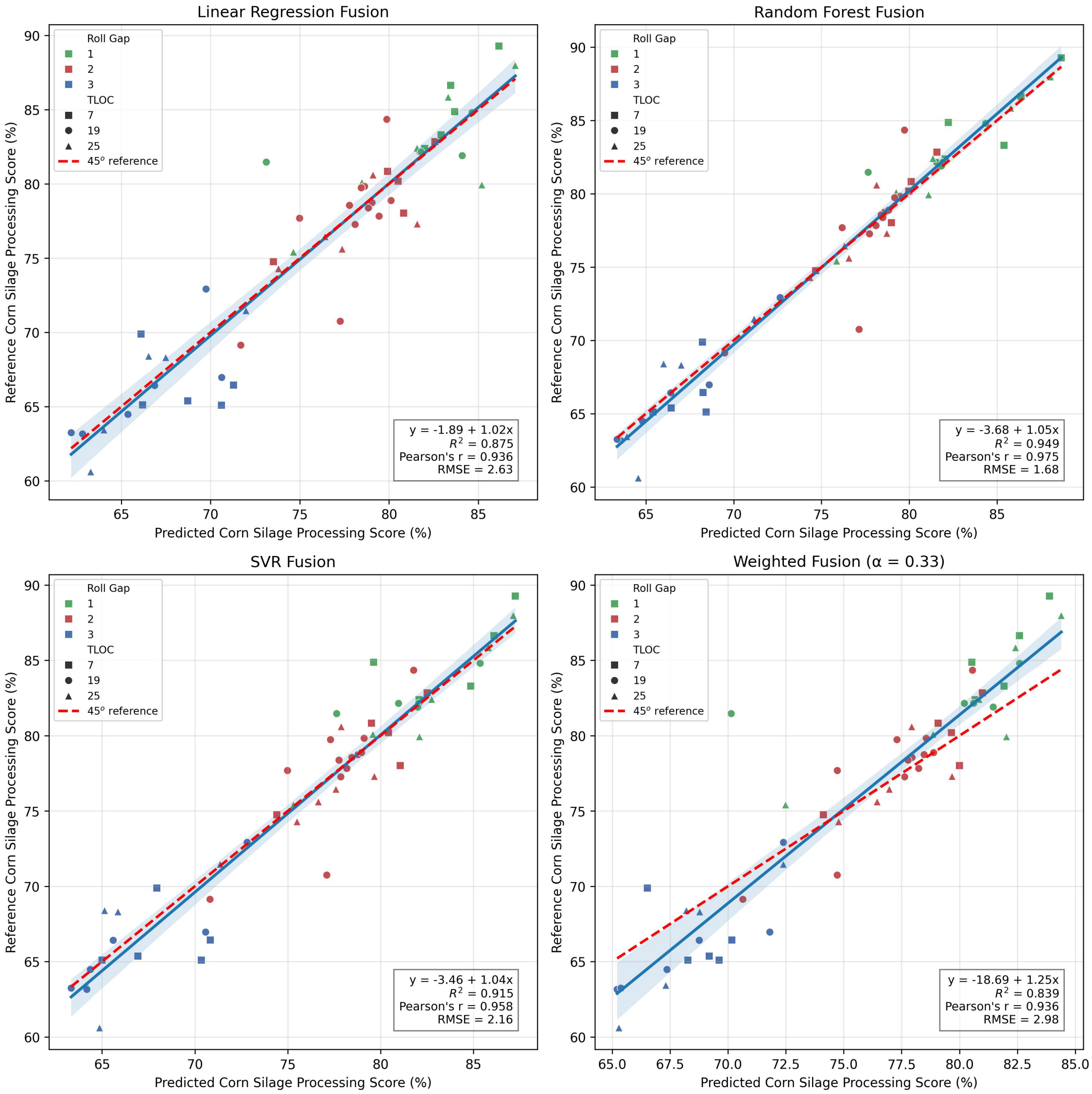


Figure 4.16: Performance of four late fusion strategies for CSPS prediction using combined RGB (FoG-based) and NIR (direct CSPS) predictions: **Top-left:** Linear Regression Fusion, **Top-right:** Random Forest Fusion, **Bottom-left:** SVR Fusion, **Bottom-right:** Weighted Fusion with $\alpha = 0.33$. Each plot compares predicted and reference CSPS values. Colors show Roll Gap levels (green = 1, red = 2, blue = 3); shapes indicate TLOC (square = 7 mm, circle = 19 mm, triangle = 25 mm). The blue line shows the regression fit with a shaded 95% confidence band. The red dashed line denotes the 1:1 ideal prediction. Random forest fusion achieved the best performance with $R^2 = 0.949$ and $RMSE_{CV} = 1.68\%$, while weighted fusion underperformed with $RMSE_{CV} = 2.98\%$.

and multimodal fusion strategies. Our results show that RGB-based models are effective when using the Fragment of Grain (FoG) as a proxy, yielding a predictive performance of

$R^2 = 0.727$ and RMSE = 2.61%. In contrast, NIR models performed best when directly predicting the laboratory-measured CSPS, achieving an $R^2 = 0.911$ and RMSE = 2.69%. These findings highlight a key distinction in the value of each modality: RGB imaging excels in estimating structural proxies like kernel fragmentation, while NIR captures chemical composition more relevant to digestibility and processing effectiveness.

Fusion approaches were explored to leverage complementary information from both modalities. The early fusion model, which concatenates RGB and NIR features, achieved moderate performance ($R^2 = 0.568$, RMSE = 5.90%), underperforming both unimodal models. This result suggests that naive concatenation can introduce feature redundancy or noise, limiting its effectiveness.

Late fusion strategies, on the other hand, produced superior results by modeling each modality independently and combining their outputs through regression or weighted averaging. The best late fusion model, using Random Forest regression on RGB-derived FoG and NIR-based CSPS predictions, achieved $R^2 = 0.949$ and RMSE = 1.68%, outperforming all individual models and fusion baselines. These results confirm that late fusion better captures the unique strengths of each sensor type and provides a scalable pathway for accurate CSPS prediction.

Overall, this study demonstrates that combining machine learning with multimodal sensing enables scalable and interpretable estimation of corn silage processing quality. The findings support real-time, non-destructive field assessment of forage processing and highlight opportunities for improved harvester calibration, quality monitoring, and decision support in precision agriculture.

Future work on machine learning may explore deep multimodal architectures or field validation of these models in real-time harvester environments. Also exploring mixed fusion would be another possibility that may be more efficient than early fusion, as we can use intermediate predictions such as FoG from RGB models to predict the CSPS.

References

[1] A. T. Badaró et al. “Identification of fiber added to semolina by near infrared (NIR) spectral techniques”. en. In: *Food Chem.* 289 (Aug. 2019), pp. 195–203.

[3] L. Breiman. “Random forests”. In: *Machine learning* 45 (2001), pp. 5–32.

[4] G. Chandrashekar and F. Sahin. “A survey on feature selection methods”. In: *Computers & electrical engineering* 40.1 (2014), pp. 16–28.

[5] T. Chen and C. Guestrin. “Xgboost: A scalable tree boosting system”. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

[6] M. Digman, K. Shinners, and T. Lemcke. “IMPACT OF FORAGE HARVESTER SETTINGS ON WHOLE-PLANT CORN PHYSICAL PROPERTIES AND CORN SILAGE PROCESSING SCORE”. In: *Transactions of the ASABE* (2025).

[8] M. Farrés et al. “Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation”. In: *Journal of Chemometrics* 29.10 (2015), pp. 528–536.

[9] L. F. Ferraretto, R. D. Shaver, and B. D. Luck. “Silage review: Recent advances and future technologies for whole-plant and fractionated corn silage harvesting”. In: *Journal of Dairy Science* 101.5 (2018), pp. 3937–3951. DOI: `10.3168/jds.2017-13728`.

[11] Y. He, X. Li, and X. Deng. “Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model”. In: *Journal of food Engineering* 79.4 (2007), pp. 1238–1242.

[12] M. A. Hearst et al. “Support vector machines”. In: *IEEE Intelligent Systems and their applications* 13.4 (1998), pp. 18–28.

[13] L. Hu et al. “Rapid detection of three quality parameters and classification of wine based on Vis-NIR spectroscopy with wavelength selection by ACO and CARS algorithms”. In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 205 (2018), pp. 574–581.

[14] H. Jiang et al. “Identification of solid state fermentation degree with FT-NIR spectroscopy: Comparison of wavelength variable selection methods of CARS and SCARS”. In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 149 (2015), pp. 1–7.

[15] W. Jiang et al. “Moving-Window-Improved Monte Carlo Uninformative Variable Elimination Combining Successive Projections Algorithm for Near-Infrared Spectroscopy (NIRS)”. In: *Journal of Spectroscopy* 2020.1 (2020), p. 3590301.

[16] K. Kawamura et al. “Vis-NIR spectroscopy and PLS regression with waveband selection for estimating the total C and N of paddy soils in Madagascar”. In: *Remote Sensing* 9.10 (2017), p. 1081.

[18] J. Li et al. “Nondestructive firmness measurement of the multiple cultivars of pears by Vis-NIR spectroscopy coupled with multivariate calibration analysis and MC-UVE-SPA method”. In: *Infrared Physics & Technology* 104 (2020), p. 103154.

[19] T. Liu et al. “NIRS feature extraction based on deep auto-encoder neural network”. In: *Infrared Physics & Technology* 87 (2017), pp. 124–128.

[21] A. Oussama et al. “Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores”. In: *Journal of the American Oil Chemists’ Society* 89 (2012), pp. 1807–1812.

[22] C. B. Rasmussen, K. Kirk, and T. B. Moeslund. “The challenge of data annotation in deep learning-A case study on Whole Plant Corn Silage”. en. In: *Sensors (Basel)* 22.4 (Feb. 2022), p. 1596.

[23] C. B. Rasmussen and T. B. Moeslund. “Evaluation of model selection for kernel fragment recognition in corn silage”. In: *arXiv [cs.CV]* (Apr. 2020).

[26] A. Rzhetsky and M. Nei. “Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference”. In: *Journal of molecular evolution* 35 (1992), pp. 367–375.

[28] S. F. C. Soares et al. “The successive projections algorithm”. In: *TrAC Trends in Analytical Chemistry* 42 (2013), pp. 84–98.

[30] B. Venkatesh and J. Anuradha. “A review of feature selection and its methods”. In: *Cybern. Inf. Technol* 19.1 (2019), pp. 3–26.

[31] W. Wang, D. Tran, and M. Feiszli. “What makes training multi-modal classification networks hard?” In: *arXiv [cs.CV]* (May 2019).

[32] C. K. I. Williams and C. E. Rasmussen. “Gaussian processes for regression”. In: *Neural Inf Process Syst* 8 (Nov. 1995), pp. 514–520.

[33] Z. Xiaobo et al. “Variables selection methods in near-infrared spectroscopy”. In: *Analytica chimica acta* 667.1-2 (2010), pp. 14–32.

Chapter 5

Combining Self-Supervised Visual Features and NIR Spectra for Optical Estimation of Corn Silage Processing Quality

William Yamada, Troy Runge, Tess Lemcke, Kevin Shinners and Matthew Digman

Publication Note: This chapter presents original research that is *intended for submission to Computers and Electronics in Agriculture.* The author of this dissertation led the project conception, implementation of self-supervised learning pipelines, feature fusion strategy, and manuscript preparation.

Abstract: Accurate estimation of the Corn Silage Processing Score (CSPS) during harvest is critical for ensuring forage quality and improving livestock feed efficiency. Traditional assessment methods are labor-intensive, require laboratory processing, and cannot provide real-time feedback. Optical approaches have emerged as alternatives, but many require extensive manual annotation, which introduces subjectivity and limits scalability. This study presents a multimodal learning framework that fuses RGB and Near-Infrared (NIR) data for CSPS prediction without the need for manual particle segmentation. RGB images were processed using self-supervised DINOv2 features, while NIR spectra were analyzed via Partial Least Squares Regression (PLSR) with Variable Importance in Projection (VIP) feature selection. We evaluated unimodal and late fusion strategies using 5-fold cross-validation across 60 field plots. The late fusion approach achieved the highest regression performance with $R^2 = 0.922$ and RMSE = 2.16%. For classification into ”Good” and ”Optimal” categories, late fusion achieved a F1 score of 0.966. Our results demonstrate that morphological and compositional features provide complementary information, and that late fusion improves predictive accuracy over unimodal models. Furthermore, the proposed method eliminates the need for manual annotations, reducing labeling errors commonly reported in the literature. Computational analysis indicates that the system can feasibly operate in near real-time on edge devices, supporting deployment during forage harvest. This work highlights the potential of self-supervised features and multimodal fusion for robust, scalable, and automated CSPS monitoring.

5.1 Introduction

Corn Silage Processing Score (CSPS) is a critical metric used to evaluate the effectiveness of kernel fragmentation during forage harvest [6]. Higher CSPS values are associated with improved digestibility, better feed efficiency, and increased milk production, making it a key target for researchers and industry practitioners [5]. Traditional laboratory methods for determining CSPS are accurate, but inherently slow, labor-intensive, and impractical for providing real-time feedback during harvesting operations. Therefore, there is growing interest in optical and machine learning-based approaches that can offer faster on-the-go quality assessments, allowing dynamic optimization of harvester settings and improving overall forage quality [4, 17, 19].

However, capturing the complexity of kernel fragmentation from optical data remains challenging due to the heterogeneous nature of silage material [16]. Variations in lighting, occlusion, and particle size distribution complicate robust and scalable assessments. These challenges have led to the exploration of computer vision and machine learning techniques to automate silage quality evaluation in field conditions.

Initial attempts to automate CSPS estimation focused on image processing techniques. For example, Drewry et al. [4] developed an algorithm to measure particle size distributions from images, achieving moderate correlation with laboratory-derived Kernel Processing Score (KPS) values. Rocha et al. [19] advanced this work by deploying a real-time imaging system on forage harvesters and using machine learning to predict KPS with high correlation to lab references.

More recent studies have explored deep learning methods for kernel detection without requiring physical separation of kernels and stover. Rasmussen and Moeslund [17] demonstrated promising results using these techniques, yet most approaches still rely heavily on large, manually annotated datasets. As noted by Rasmussen et al. [16], manual annotation in silage images is highly laborious, error-prone, and inconsistent—particularly when occlusions and small particles obscure kernel boundaries. These issues hinder scalability and reduce model generalizability.

To alleviate the annotation bottleneck, self-supervised learning (SSL) has gained attention for its ability to learn representations from unlabeled data. In agriculture, SSL has been shown to improve performance in classification tasks with minimal annotations. Kar et al. [8] reported significant improvements in classifying agriculturally important insect pests using SSL, and Güldenring and Nalpantidis [7] demonstrated superior plant classification performance from contrastive SSL pre-training.

Building on these developments, Sornapudi and Singh [20] proposed a lightweight SimCLR-based framework for agricultural vision tasks, while Marszalek et al. [12] highlighted SSL's potential in precision agriculture for reducing annotation effort. Surveys and frameworks by Li [9], Pinto-da-Silva [15], and Wang [21] further reinforce the utility of foundation models and SSL for handling label scarcity, domain shifts, and multimodal sensing.

In light of these trends, this study aims to develop an optical CSPS estimation framework that:

- avoids the need for manual segmentation annotations by leveraging self-supervised features extracted from DINOv2, a vision transformer pretrained with self-distillation;
- combines RGB-derived morphological information with NIR-derived compositional information through late fusion to improve prediction accuracy and robustness;
- evaluates regression performance for CSPS estimation as well as classification performance to distinguish between “Good” (50–70%) and “Optimal” (¿70%) CSPS quality classes;
- and assesses the feasibility of deploying the resulting models on resource-constrained, edge-computing platforms suitable for real-time, in-field operation.

By addressing the dual challenges of annotation cost and generalization, this work contributes toward scalable, deployable solutions for real-time silage quality monitoring in modern agricultural systems.

5.2 Materials and Methods

This study explores unimodal and multimodal approaches to estimate the Corn Silage Processing Score (CSPS), a key indicator of forage quality. The goal was to evaluate predictive performance when using either RGB images or near-infrared (NIR) spectra independently, and to assess whether a late-fusion strategy combining both modalities could yield improved CSPS estimation and classification results.

Field data were collected during the fall of 2021 at the Arlington Research Station in Wisconsin, USA (43.303°N, 89.346°W). For each plot, a John Deere FQ Camera captured high-resolution RGB images at 0.068 mm/pixel, generating 30 images per plot. Each image was divided into three non-overlapping crops of 468×468 pixels, promoting diversity while maintaining complete spatial coverage. Concurrently, a John Deere HarvestLab 3000 spectrometer acquired NIR reflectance spectra from 950 to 1650nm using 256 diodes. At least five NIR scans were collected per plot to ensure robustness.

Ground truth CSPS values were obtained via laboratory analysis following standard protocols Ferreira2005-fz. In addition, the fraction of material retained above a 4.75 mm sieve—denoted as *FoG ¿ 4.75%*—was recorded as a proxy for kernel processing efficiency, allowing for interpretable modeling.

For RGB-based modeling, image features were extracted using DINOv2 oquab2023dinov2, a self-supervised Vision Transformer (ViT) model developed by Meta AI. DINOv2 achieves state-of-the-art performance across a wide range of computer vision tasks without requiring manual labels, making it particularly suitable for applications where data annotation is difficult or impractical [2, 14]. Its architecture is well suited for capturing complex visual patterns, which is essential when analyzing the heterogeneous and cluttered appearance of silage images. Additionally, Darcet et al. [3] demonstrated that incorporating explicit register tokens in ViTs improves both the interpretability of attention maps and downstream task performance. For each image crop, the latent representations of all patches were extracted and averaged to obtain a plot-level feature vector.

Two regression strategies were tested using these features: one directly predicted CSPS using a Random Forest regressor, while the other first predicted the percentage of coarse particles retained on the 4.75 mm sieve (FoG) and transformed the predicted value into CSPS via:

$$\widehat{\text{CSPS}}_{\text{RGB}} = 99.24 - 0.93 \cdot \widehat{\text{FoG}} \tag{5.1}$$

This equation was derived empirically and demonstrated strong predictive validity ($R^2 = 0.853$, $p \ll 0.05$).

For NIR-based modeling, spectra were first preprocessed using Standard Normal Variate (SNV) normalization, followed by a Savitzky–Golay filter (window size of 11, polynomial order 2, first derivative). Variable Importance in Projection (VIP) scores were used to select relevant wavelengths, and Partial Least Squares Regression (PLSR) was trained to estimate CSPS.

A late fusion framework was developed to combine predictions from the RGB and NIR pipelines. Linear regression and Random Forest regression were evaluated as meta-models. Each base model was trained under a five-fold cross-validation protocol, and predictions from held-out folds were used as inputs to the fusion model. An overview of this multimodal workflow is provided in Figure 5.1.

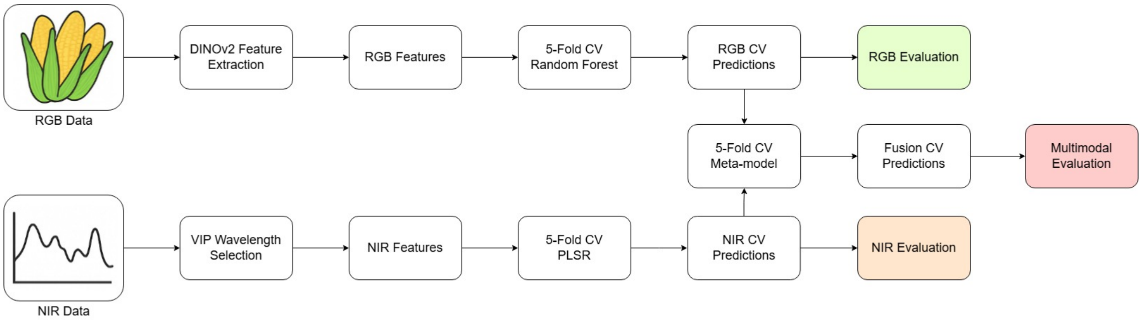


Figure 5.1: Overview of the RGB and NIR unimodal pipelines and late fusion strategy. RGB features are extracted using DINOv2 and evaluated via Random Forest. NIR spectra undergo VIP-based wavelength selection followed by PLSR. A meta-model combines 5-fold predictions from both pipelines.

Both regression and binary classification tasks were investigated for each modality and their fusion. For classification, CSPS values were thresholded into two quality classes: *Good* (50–70%) and *Optimal* (¿70%). Models were evaluated using 5-fold stratified cross-validation to maintain balanced class distributions and prevent plot-level data leakage.

Regression performance was assessed using Root Mean Squared Error ($RMSE_{CV}$), Coefficient of Determination (R^2), and Pearson correlation coefficient (r). Together, these metrics provide complementary insights into model accuracy, variance capture, and trend alignment.

Classification performance was evaluated using Accuracy, F1-score, ROC AUC, and confusion matrix components (TP, FP, TN, FN), enabling both threshold-dependent and threshold-independent assessments of model behavior.

All analyses were implemented in Python 3.10.12 using `scikit-learn`, `scipy`, `numpy`, and `matplotlib`. Feature extraction from DINOv2 was conducted using PyTorch and the

Hugging Face Transformers library.

5.3 Results and Discussion

5.3.1 DINOv2 Attention Head Analysis

Before proceeding to the dataset feature extraction, we performed a preliminary evaluation of the DINOv2 feature extractor to assess whether the foundational model captured salient information related to kernel structures in whole-plant corn silage images. Specifically, we inspected the outputs from all 16 attention heads of the final DINOv2 transformer block across a sample dataset.

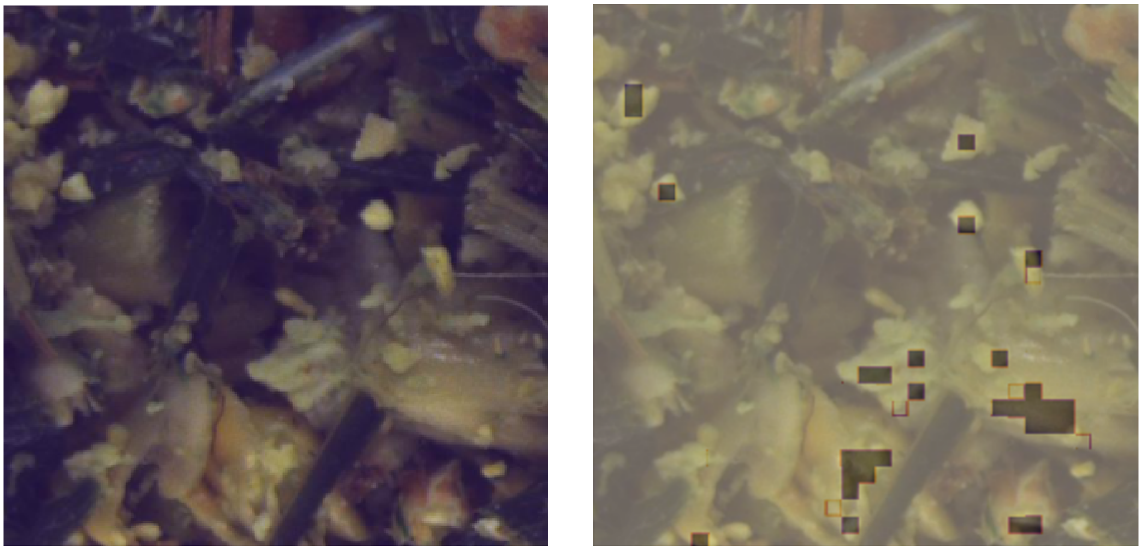
As illustrated in Figure 5.2, some attention heads appeared capable of highlighting kernel fragments and coarse material without any supervised fine-tuning. Notably, one head consistently responded to regions resembling corn kernels, even under the challenging imaging conditions characterized by occlusion, clutter, and varying particle sizes. This observation reinforced the suitability of DINOv2 as the RGB feature extractor, suggesting that self-supervised representations from a generalist model could capture biologically relevant structures for downstream tasks.

5.3.2 Regression Performance

To evaluate the regression performance of each modality and fusion strategy for CSPS prediction, we conducted 5-fold cross-validation and report R^2, Pearson correlation, and $RMSE_{CV}$ across models. Both unimodal and multimodal approaches are compared.

Table 5.1: Summary of regression performance across models.

Model	R^2	**Pearson's** r	$RMSE_{CV}$
RGB → CSPS (direct)	0.423	0.676	5.85 %
RGB → FoG → CSPS	0.444	0.717	5.75 %
NIR → CSPS (direct)	0.425	0.695	5.84 %
NIR → FoG → CSPS	0.353	0.662	6.79 %
Late Fusion (Linear)	0.604	0.777	4.85 %
Late Fusion (RF)	**0.922**	**0.968**	**2.16 %**



(a) Original RGB sample image

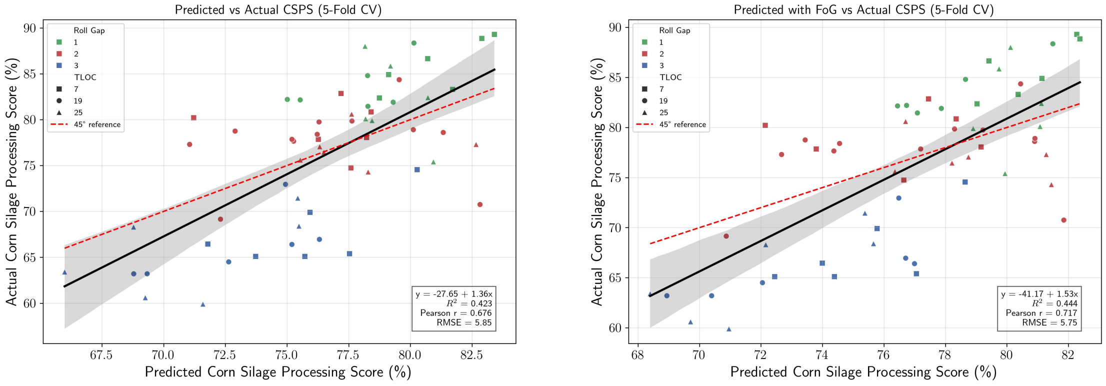
(b) Highlighted attention regions (from one DINOv2 head)

Figure 5.2: Preliminary assessment of DINOv2 attention heads. Several heads captured features resembling kernel fragments without supervised training.

As shown in Table 5.1 and Figure 5.3, late fusion using Random Forest outperformed all other models with an R^2 of 0.922 and $RMSE_{CV}$ of 2.16 %. This suggests that RGB and NIR modalities provide complementary information for predicting CSPS. The unimodal NIR model and RGB-based proxy also performed reasonably well (R^2 around 0.44), indicating that morphological or spectral cues alone can partially explain CSPS variability, but not to the same extent as fused models.

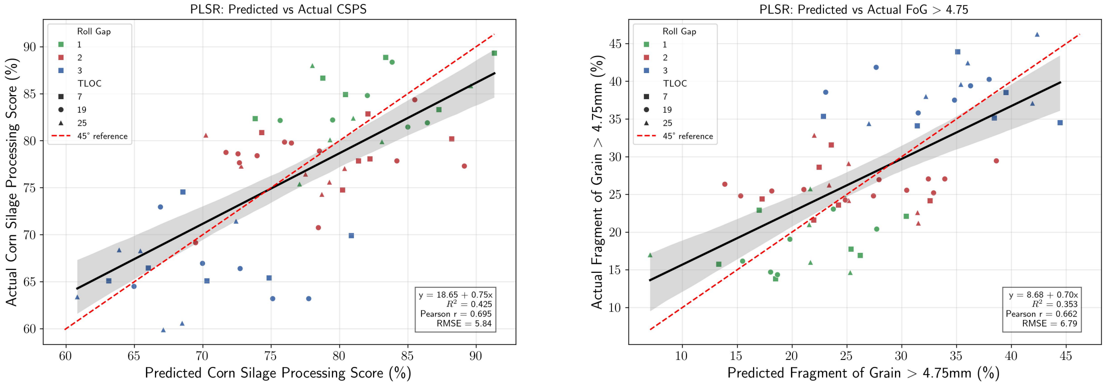
The intermediate use of FoG (fraction of particles > 4.75 mm) as a proxy for estimating CSPS proved beneficial, particularly when derived from RGB features. Although indirect, the FoG-based pipeline improved prediction metrics compared to direct RGB-CSPS regression, supporting the relevance of FoG as a proxy variable. This proxy may be especially useful in practical applications where direct CSPS ground truth is unavailable.

The best fusion results were achieved by training on predicted CSPS values from RGB features using the FoG proxy, combined with direct CSPS estimates from NIR spectra selected by PLSR-VIP. The optimal linear combination was achieved using a fusion weight of $\alpha = 0.61$, giving slightly more influence to the RGB-derived proxy.



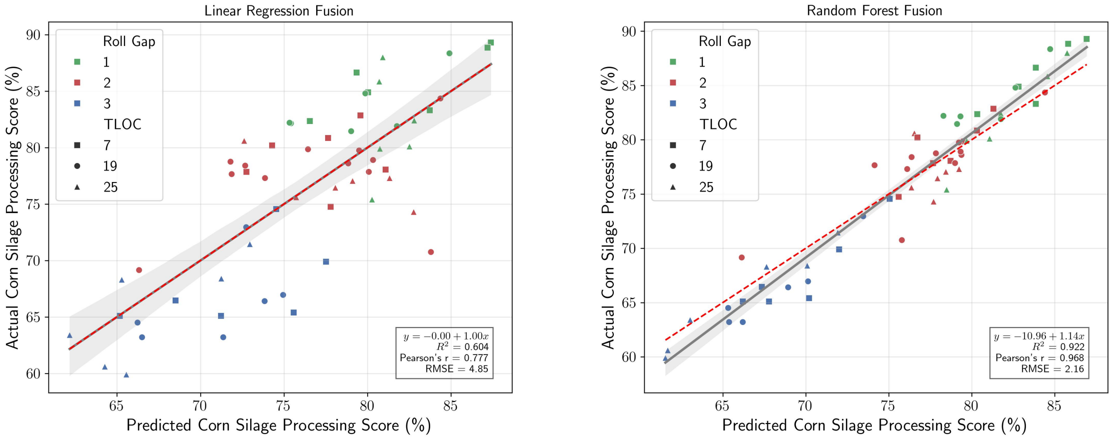
(a) RGB → CSPS (direct regression)

(b) RGB → FoG → CSPS



(c) NIR→ CSPS (direct regression)

(d) NIR→ FoG → CSPS



(e) Late Fusion (Linear Regression)

(f) Late Fusion (Random Forest)

Figure 5.3: Comparison of regression results across all models for CSPS prediction. Each point represents a plot-level prediction using 5-fold cross-validation. Colors indicate Roll Gap (green = 1, red = 2, blue = 3), and shapes represent TLOC (square = 7 mm, circle = 19 mm, triangle = 25 mm). The solid gray line shows the regression fit with 95% confidence band; the dashed red line denotes the 1:1 ideal reference.

Most prediction errors occurred in plots with extreme CSPS values, particularly those close to the decision boundary between good and suboptimal CSPS values. These errors suggest that the models are sensitive to moderate processing quality levels but less accurate under conditions of excessive processing or grain retention. To fully evaluate this observation, more data with lower processing quality will be required.

The predictive accuracy of the fused model, especially using Random Forests, suggests the feasibility of deploying this system for real-time CSPS monitoring during harvest. Given the use of compact RGB imaging and miniaturized NIR sensors, the models presented here could be implemented onboard forage harvesters, enabling precision adjustments to processing equipment. Additionally, the DINO feature extractor for RGB imagery, despite its large size for training, can be distilled or frozen for lightweight inference pipelines. Combined with the low inference time and robustness of Random Forest regressors, this approach would allow onboard computation without requiring cloud-based processing, thereby supporting real-time decision-making in the field.

When comparing the results of this study to prior literature, it is evident that the proposed fusion approach offers significant advancements. Drewry et al. [4] developed an image processing algorithm using particle size metrics to predict Kernel Processing Score (KPS), achieving a Pearson correlation of $r = 0.80$ relative to mechanical sieving. Their method relied solely on 2D imaging within a controlled environment and did not incorporate spectral information. Rocha et al. [19] improved upon this by introducing a high-speed imaging system combined with machine learning kernel detection, achieving a very high correlation ($r = 0.977$) and a standard error of 2.71% for KPS estimation, although their pipeline included counting whole kernels, which might be a rare occasion on well calibrated processors.

Rasmussen and Moeslund [17] explored deep learning-based kernel detection without kernel/stover separation, reporting a strong correlation ($r = 0.88$) between model-derived and optical KPS values, although the segmentation precision remained moderate (average precision $\approx$ 34–36%). Their work highlights the challenges of achieving accurate kernel estimation in field conditions without separation steps. The missing link between Optical CSPS and laboratory reference CSPS was the main gap in this study.

In contrast, our late fusion model achieved a Pearson's $r = 0.968$ and $RMSE_{CV}$ of 2.16 when predicting CSPS while relying on combined RGB morphological features and NIR compositional data. Moreover, the proposed pipeline uses compact and feasible components for field deployment, with the DINO feature extractor providing strong generalization from RGB images and Random Forest regressors offering low-latency inference. Therefore, this study presents a practical, lightweight alternative capable of operating in real-time onboard harvesters, expanding the possibilities for in situ quality assessment of corn silage beyond kernel-centric metrics alone.

To assess deployment feasibility, we estimate that the late fusion system can achieve real-time performance on modern edge devices. Optimized DINO variants, such as Grounding DINO 1.5 Edge, have demonstrated over 10 FPS on the NVIDIA Orin NX using TensorRT optimization [18]. Random Forest regressors also offer low inference overhead, with approaches like Bolt achieving 2× faster predictions compared to previous methods [13]. Together, these components enable CSPS estimation within 100 milliseconds per sample on platforms like the NVIDIA Jetson series, supporting onboard, real-time operation during forage harvesting. However, given the limited dataset of 60 plots from a single season, future work should address generalization across years and locations.

5.3.3 Classification Performance

Classification analysis was conducted to categorize CSPS values into “Good” (50–70%) and “Optimal” (>70%) classes using RGB-only, NIR-only, and Late Fusion strategies. Key metrics, confusion matrices, and ROC/PR curves were evaluated.

Table 5.2 highlights the classification performance for each modeling strategy. And each model will be explored individually:

RGB-only Classification: The RGB-based classifier achieved an AUC of 0.911, precision of 0.860, and recall of 0.977 at the optimal threshold (0.40). As shown in

Table 5.2: Classification metrics summary for RGB-only, NIR-only, and Late Fusion models. The P column stand for Precision, the R column stand for recall.

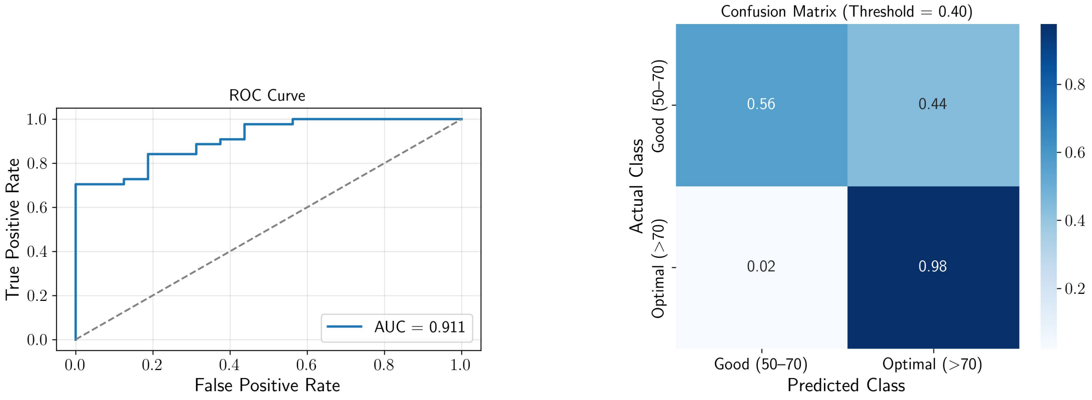
Model	**P**	**R**	**F1**	**ROC AUC**
RGB-only	0.860	0.977	0.915	0.911
NIR-only	0.530	1.000	0.696	0.559
Late Fusion	0.977	0.955	0.966	0.994

Figure 5.4, the feature space exhibited clear clustering between the two classes. The confusion matrix reveals a strong ability to correctly classify Optimal samples, with minimal false negatives. However, some Good samples were misclassified as Optimal, suggesting a slight overprediction bias likely linked to the proxy estimation step (FoG → CSPS). The F1 score peaked at 0.915, indicating robust overall performance.

NIR-only Classification: The NIR-based classifier performed substantially worse, achieving an AUC of 0.559 and F1 score of 0.696 at the best threshold (0.17). Figure 5.5 shows that the classifier had difficulty separating Good and Optimal classes, with many false positives for Optimal samples. Precision was high (1.0) for Good samples but recall was very low (0.12), suggesting that the NIR model struggled to identify Good-quality processing. The PR curve was flat and ROC close to the diagonal, confirming the limited discriminative ability when using spectral features alone for this task.

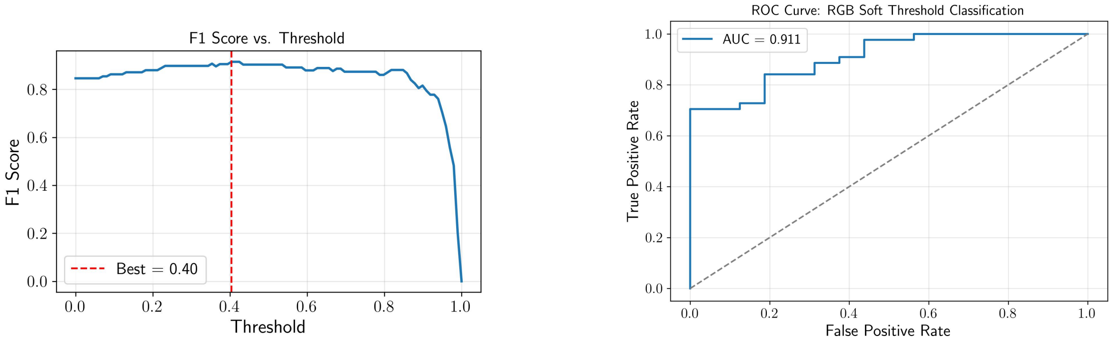
Late Fusion Classification: Late fusion greatly improved classification performance across all metrics. As shown in Figure 5.6, the model achieved an AUC of 0.994, a precision of 0.977, recall of 0.955, and F1 score of 0.966. The feature space exhibited almost perfect linear separability. The confusion matrix confirmed high classification accuracy, with very few false positives and false negatives. The PR curve was nearly ideal (AP = 0.998), and the ROC curve showed a steep ascent with minimal area under the diagonal. The fusion leveraged complementary strengths from both RGB (morphology) and NIR (compositional quality), overcoming the individual weaknesses of the unimodal approaches.

Overall, the classification results mirror the regression findings: while RGB-only and NIR-only models provide partial predictive power, combining modalities through late fusion substantially improves reliability. These results reinforce the growing evidence that late



(a) Feature space for RGB-based CSPS classification. Each point represents a sample colored by the ground truth class (Good = blue, Optimal = red). A clear horizontal separation can be observed, indicating that RGB-derived features (FoG-based) were informative for distinguishing processing quality classes.

(b) Confusion matrix for RGB-based classification at threshold 0.40. The model exhibited high true positive rates for Optimal class predictions, with a few Good samples misclassified as Optimal, indicating a slight tendency toward overestimation.

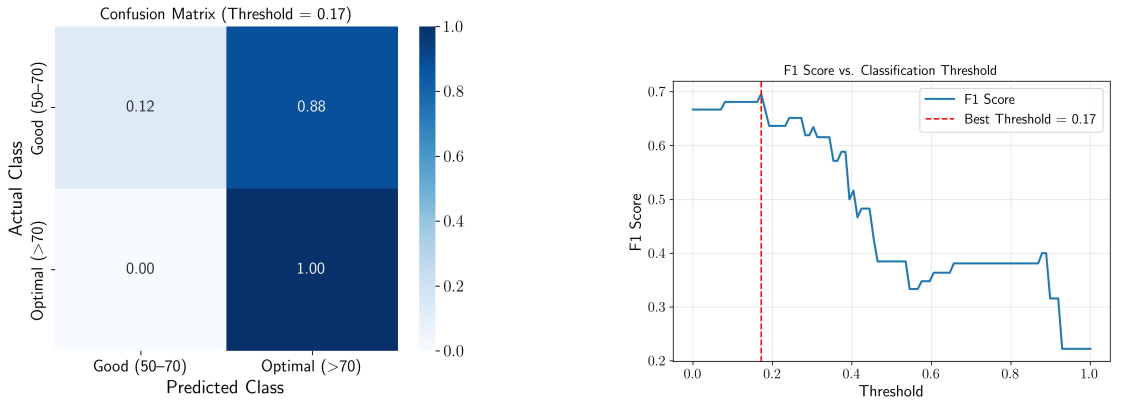


(c) F1 score as a function of threshold for RGB-based classification. The optimal threshold maximizing the F1 score was identified at 0.40, balancing precision and recall effectively for the two-class prediction task.

(d) Receiver Operating Characteristic (ROC) curve for RGB-based classification, achieving an AUC of 0.911. The curve indicates strong discriminatory ability, although some sensitivity loss is observed at very low false positive rates.

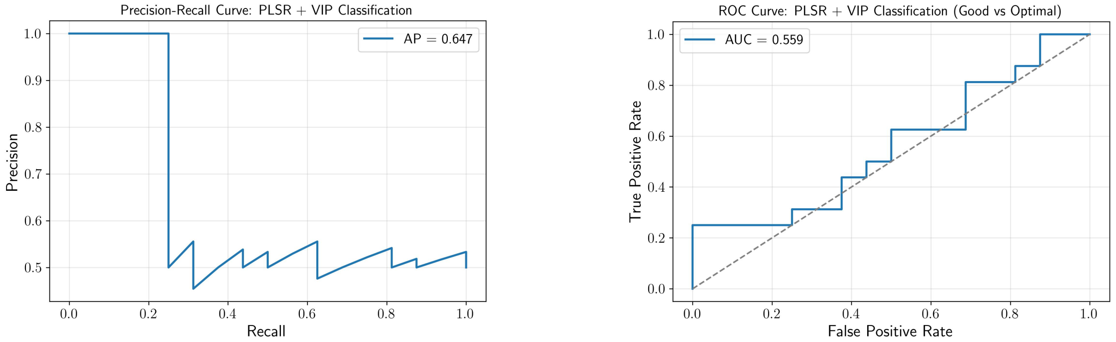
Figure 5.4: Classification results using only RGB-derived features for CSPS threshold prediction. Despite good separability, minor misclassifications suggest limitations when relying solely on RGB morphology-based proxies.

fusion is preferable for small-to-medium datasets where early feature-level fusion may overfit or fail to properly align heterogeneous signals [1, 10].



(a) NIR Confusion Matrix

(b) NIR F1 vs Threshold



(c) NIR Precision-Recall Curve

(d) NIR ROC Curve

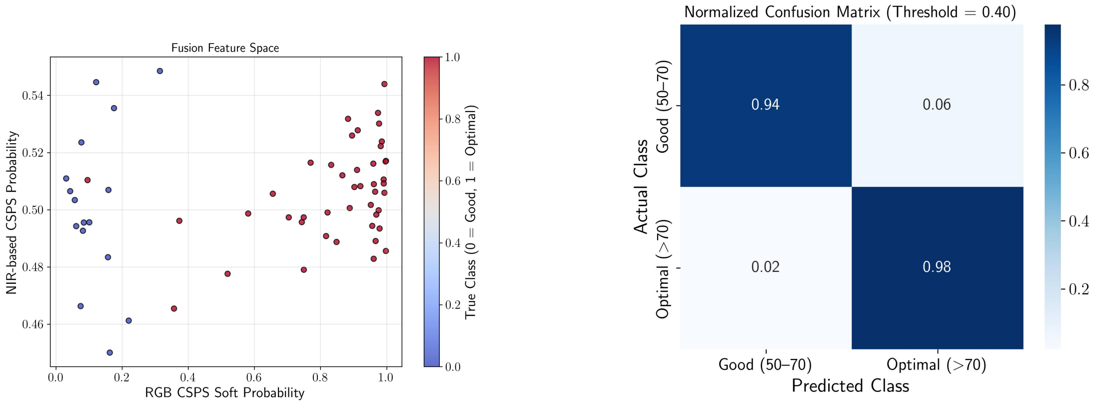
Figure 5.5: NIR-only classification results for CSPS thresholding.

5.4 Comparison with Existing Literature

Prior studies have explored a variety of techniques for estimating kernel processing scores (KPS) and corn silage processing score (CSPS), ranging from traditional image processing pipelines to deep learning and real-time machine-integrated systems. Table 5.3 summarizes key findings and compares them with the current work.

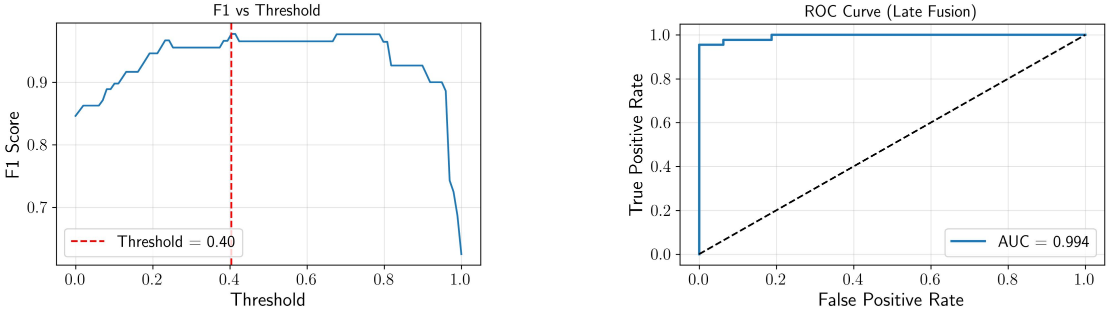
Compared to earlier image-only methods [4, 11], the proposed multimodal approach benefits from integrating both morphological (RGB) and compositional (NIR) information. Whereas previous works relied on geometric proxies or kernel counting, this study eliminates the need for manual annotation or explicit object detection by leveraging self-supervised features extracted from DINOv2.

Rocha et al. [19] achieved excellent correlation ($r = 0.977$) using high-speed imagery



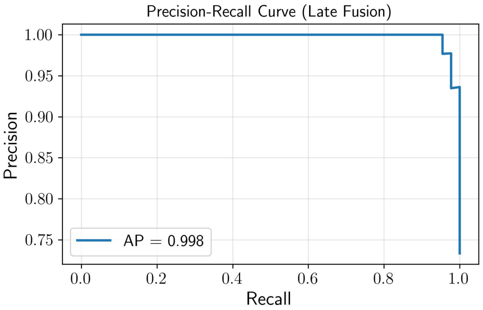
(a) Late Fusion Feature Space

(b) Late Fusion Confusion Matrix



(c) Late Fusion F1 vs Threshold

(d) Late Fusion ROC Curve



(e) Late Fusion Precision-Recall Curve

Figure 5.6: Late fusion classification results combining RGB and NIR.

and kernel counting, but their method requires detecting whole kernels—a rare condition in properly calibrated harvesters. Rasmussen and Moeslund [17] explored deep object recognition for kernels but reported only moderate segmentation precision (AP $\approx$ 34–36%) and relied heavily on detailed labels.

This study improves upon prior work by:

Table 5.3: Comparison of prior CSPS/KPS estimation methods with the proposed multimodal approach.

Study	**Method**	**Highlights**	**Best r or R^2**	**Limitations**
Drewry et al. (2019) [4]	Image processing of kernel particles	Estimated KPS using geometric features (e.g., MICD, area); $r = 0.80$ with sieving reference	$r = 0.80$	Lab-based, no NIR
Luck et al. (2020) [11]	Smartphone image processing (SilageSnap)	Correlated image-based particle size to in-situ digestibility; field-ready app	$r = 0.77$	No spectral data, indirect validation
Rocha et al. (2022) [19]	ML + image kernel counting (spout camera)	High-speed image acquisition + ML kernel counting; offline linear regression for KPS	$r = 0.977$	Kernel counting limits generalization; image-only
Rasmussen and Moeslund (2019) [17]	Deep learning object detection	Segmentation of kernels in RGB; AP ≈ 34–36%; $r = 0.88$ KPS correlation	$r = 0.88$	Relied on pixel-wise labels, moderate precision
This study (2025)	**RGB+DINOv2 + NIR+VIP + late fusion**	**No manual segmentation; self-supervised features + spectral data; deployed fusion model**	$r = 0.968$, $R^2 = 0.922$	Limited to single season, 60 plots

- Utilizing self-supervised RGB features (DINOv2) that require no labeled images;
- Combining RGB and NIR modalities via late fusion to enhance robustness;
- Achieving high predictive accuracy ($r = 0.968$, $R^2 = 0.922$, $RMSE_{CV} = 2.16\%$) without kernel segmentation;
- Designing a pipeline suitable for real-time deployment using edge computing.

These advances suggest that multimodal learning provides a scalable, interpretable, and high-performing alternative to traditional image-only or handcrafted methods for silage processing quality estimation.

5.5 Study Limitations

Despite the strong performance of the proposed multimodal framework, several limitations should be acknowledged:

- **Dataset size and diversity:** The dataset used in this study consisted of only 60 plots, all collected during a single harvest season at a single location. While stratified cross-validation was used to improve generalizability, this limited sample size restricts the external validity of the model. Future work should incorporate multi-season, multi-location datasets to ensure robustness across varying agronomic conditions, hybrid varieties, and environmental factors.
- **Range of CSPS values:** The dataset was skewed toward well-processed silage, with relatively few samples exhibiting very low CSPS values. As a result, model performance may degrade under suboptimal or edge-case processing conditions. More balanced sampling across the entire spectrum of processing quality is required to build models capable of robust predictions in operational extremes.
- **Dependence on proxy variables:** While the use of the Fragment of Grain (FoG) as an intermediate variable improved performance, it also introduces an additional layer of approximation. The effectiveness of this proxy may vary depending on the crop, harvest equipment, or particle visibility in images, potentially limiting generalization.
- **Sensor alignment and synchronization:** Although RGB and NIR data were collected in close proximity on the harvester spout, perfect spatial and temporal alignment between modalities was not guaranteed. Small mismatches could introduce noise into the fusion process. Integrated multi-sensor calibration and alignment strategies could improve data fusion fidelity.
- **Edge deployment trade-offs:** While Random Forests were chosen for their favorable inference speed and DINOv2 features were extracted offline, deploying the entire pipeline in real-time on embedded systems may require additional optimization such

as model quantization or distillation. Real-world validation on edge hardware was estimated but not executed in this study.

- **Lack of external validation:** Model evaluation was limited to internal cross-validation folds. Although this approach reduces overfitting risk, an independent hold-out dataset or external test site validation is essential for rigorous benchmarking before field deployment.

These limitations highlight important considerations for future work and motivate continued development toward generalizable, scalable, and interpretable systems for real-time silage quality monitoring.

5.6 Conclusion

This study demonstrated the effectiveness of combining morphological (RGB) and compositional (NIR) features for predicting corn silage processing scores (CSPS). Through a comprehensive comparison of unimodal and multimodal approaches, we showed that late fusion strategies significantly outperformed single-modality models, achieving an R^2 of 0.922 for regression and an F1 score of 0.966 for classification.

Regression results indicated that while both RGB-derived FoG proxies and NIR-derived spectral features provided moderate predictive ability individually, fusing their complementary information through a Random Forest model yielded the most accurate CSPS estimations. The optimal fusion strategy placed greater weight on RGB-based morphological cues ($\alpha = 0.61$), while still leveraging the spectral sensitivity of NIR data.

Classification results further reinforced the benefits of multimodal fusion. The late fusion model exhibited the highest precision, recall, and F1-score among the tested approaches, substantially reducing false positives and false negatives compared to unimodal models. These outcomes highlight the potential of integrating both RGB imaging and NIR spectroscopy for real-time silage quality assessment.

A key advantage of the proposed system is that it did not require manual segmentation annotations of kernels or particles. Instead, feature extraction relied on self-supervised learning and regression targets based on laboratory-measured CSPS values, avoiding the inconsistencies and labeling errors often associated with manual annotation processes [16]. This makes the approach substantially more scalable and reproducible compared to prior deep learning-based methods that depend heavily on detailed pixel-level annotations.

From a deployment perspective, the combination of frozen DINO feature extractors, lightweight Random Forest models, and embedded NIR analysis suggests a feasible path toward implementing these systems onboard forage harvesters. Real-time inference, without reliance on cloud computing, could enable farmers to monitor and adjust kernel processing quality dynamically during harvest operations.

Future work should validate the approach across broader environments, including different seasons, hybrid varieties, and harvest conditions. Additionally, interpretability analyses could provide insights into the most influential morphological and spectral features, paving the way for more explainable precision agriculture systems.

References

[1] T. Baltrušaitis, C. Ahuja, and L.-P. Morency. “Multimodal machine learning: A survey and taxonomy”. In: *arXiv [cs.LG]* (May 2017).

[2] M. Caron et al. “Emerging properties in self-supervised vision transformers”. In: *arXiv [cs.CV]* (Apr. 2021).

[3] T. Darcet et al. “Vision Transformers Need Registers”. In: *arXiv [cs.CV]* (Sept. 2023).

[7] R. Güldenring and L. Nalpantidis. “Self-supervised contrastive learning on agricultural images”. en. In: *Comput. Electron. Agric.* 191.106510 (Dec. 2021), p. 106510.

[8] S. Kar et al. “Self-supervised learning improves classification of agriculturally important insect pests in plants”. en. In: *The Plant Phenome Journal* 6.1 (Dec. 2023).

[9] J. Li et al. “Foundation models in smart agriculture: Basics, opportunities, and challenges”. en. In: *Comput. Electron. Agric.* 222.109032 (July 2024), p. 109032.

[10] X. Liu et al. “Late Fusion Incomplete Multi-view clustering”. en. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 41.10 (Oct. 2019), pp. 2410–2423.

[12] M. L. Marszalek et al. “Self-supervised learning – a way to minimize time and effort for precision agriculture?” en. In: *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLIII-B3-2022 (May 2022), pp. 1327–1333.

[13] Y. Ni et al. “Online performance and power prediction for edge TPU via comprehensive characterization”. In: *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE).* IEEE, Mar. 2022, pp. 612–615.

[14] M. Oquab et al. “DINOv2: Learning robust visual features without supervision”. In: *arXiv [cs.CV]* (Apr. 2023).

[15] M. Pinto da Silva et al. “From tradition to transformation: Deep and self-supervised learning approaches for remote sensing in agriculture and environmental change”. In: *Social Science Research Network* (Dec. 2024).

[18] T. Ren et al. “Grounding DINO 1.5: Advance the “edge” of open-set object detection”. In: *arXiv [cs.CV]* (May 2024).

[20] S. Sornapudi and R. Singh. “Self-supervised backbone framework for diverse agricultural vision tasks”. In: *arXiv [cs.CV]* (Mar. 2024).

[21] H. Wang et al. “CC-SSL: A self-supervised learning framework for crop classification with few labeled samples”. en. In: *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 15 (2022), pp. 8704–8718.

Chapter 6

Conclusion and Future Work

Conclusion

Near-Infrared (NIR) spectroscopy is rapidly becoming a cornerstone technology in the digital transformation of agriculture. Its ability to non-destructively estimate the chemical composition of biological materials—such as moisture, protein, fiber, starch, and lignin—has positioned it as a valuable tool for monitoring crop and forage quality, assessing soil health, and verifying food integrity. What once required lab-based wet chemistry and time-consuming sampling can now be achieved in seconds with portable NIR devices deployed directly in the field, on machinery, or even integrated into autonomous platforms. This evolution enables real-time, spatially resolved decisions across the production pipeline—from planting and fertilization to harvest and storage.

The practical impact of NIR is seen in its versatility: it supports ration formulation for livestock by evaluating forage digestibility, helps detect grain adulteration and contamination, and provides insights into soil nutrient dynamics—all without altering the sample. As the technology becomes more compact and affordable, its potential for democratization expands. Producers, agronomists, and food processors can now access lab-grade analytics in the field, enabling more precise and sustainable practices. While machine learning and AI have enhanced the modeling and interpretation of complex spectral data, the strength of NIR lies in its direct linkage to physical and chemical properties, providing a scientifically grounded signal that bridges traditional agronomy and modern data science. Looking ahead, the continued fusion of NIR with emerging technologies—including imaging, edge computing, and cloud-based analytics—promises to deliver scalable, explainable, and high-impact tools for the next generation of agriculture.

This dissertation has advanced the integration of spectroscopic and image-based sensing for real-time assessment of forage processing quality and crop loss in agricultural systems. Across four distinct but interconnected studies, the work demonstrated the feasibility, accuracy, and adaptability of multimodal machine learning frameworks under practical, in-field constraints.

First, handheld NIR spectrometers were evaluated for estimating nutritional traits in undried forage. Despite challenges posed by water absorption and field variability, the study showed that appropriate preprocessing and calibration methods can yield robust models for crude protein, fiber, and digestibility. This affirms the viability of portable NIR systems for in situ forage quality screening.

Second, pre-harvest loss quantification was explored using deep learning-based object detectors applied to field images of soybean, oat, and wheat losses. These models effectively localized and quantified uncollected grain, paving the way for visual feedback loops in harvest optimization.

Third, a late-fusion strategy was proposed that integrated RGB-derived morphological features with NIR spectra. This approach significantly outperformed unimodal baselines, achieving a Pearson correlation of 0.968 and RMSECV of 2.16% for CSPS prediction, and a binary classifier AUC of 0.994. The findings underscore the potential of synergistic data fusion to capture both structural and compositional traits.

Finally, self-supervised learning was applied to extract semantic features from silage images using DINOv2 Vision Transformers. These models performed strongly in estimating CSPS without requiring manual annotation or kernel segmentation, reinforcing the value of foundation models and weak supervision in agricultural machine vision. The use of a proxy variable—fraction of grain retained over 4.75 mm (FoG)—further improved explainability and interpretability.

Together, these contributions provide a scalable, interpretable framework for deploying intelligent, multimodal sensing systems in precision agriculture. The results advocate for a shift from isolated sensor modalities to unified perception architectures capable of generalizing across dynamic agricultural environments.

Beyond academic contributions, the practical applications of the systems developed in this dissertation are far-reaching. When deployed on harvesting implements, these multimodal sensing systems could enable automated control of harvest settings—dynamically adjusting roll gap or kernel processor speeds to optimize forage digestibility and minimize grain loss in real-time. Such adaptability has the potential to enhance feed efficiency, reduce input waste, and improve economic returns for producers.

Moreover, the fusion of visual and spectral data opens pathways for automated in-field phenotyping, supporting timely agronomic decisions. By identifying traits such as biomass, kernel fragmentation, or nutritional value during key growth stages, these tools can guide in-season adjustments to nutrient applications or pest management. This not only reduces the environmental footprint of farming practices but also supports data-driven yield optimization.

Importantly, the adoption of compact, affordable sensors—paired with interpretable machine learning—offers a low-cost upgrade path for existing farm equipment. By transforming tractors, harvesters, or UAVs into mobile crop laboratories, producers can routinely monitor crop quality and loss metrics without the delays and costs associated with laboratory testing. This democratization of analytics has the potential to increase access to precision agriculture technologies for small and mid-sized operations.

As agriculture continues to evolve under the pressures of sustainability, climate variability, and global food demand, the ability to generalize sensing models across crops, seasons, and geographies will be essential. The scalable framework presented in this work lays the foundation for such adaptability—blending physical insight, machine learning, and edge deployability into a unified system ready for the next generation of intelligent farming.

Future Work

While this research has addressed critical gaps, several avenues remain open for continued exploration:

- **Cross-Site and Temporal Generalization:** The current models were validated using single-season, single-site datasets. Extending to multi-year and multi-location trials is essential to evaluate model stability under varying environmental, genetic, and operational conditions.
- **Model Interpretability and Uncertainty Quantification:** Attention maps and VIP scores provided initial insight into feature relevance. Future work should explore SHAP values, Bayesian neural networks, or ensemble-based uncertainty measures to improve interpretability and trust in model outputs.
- **Embedded Deployment and Real-Time Processing:** Implementation on embedded devices such as NVIDIA Jetson modules should be pursued. This includes profiling inference latency, leveraging TensorRT acceleration, and ensuring sensor synchronization for truly on-the-go deployment.
- **Pre-Harvest Trait Prediction:** While this work focused on harvested material, the same sensing framework could be extended to early-season or mid-season traits, such as grain fill, biomass accumulation, lodging risk, or stress detection, enabling earlier intervention strategies.
- **User-Facing Interfaces and Decision Tools:** Future systems should integrate visualization dashboards and feedback loops, allowing end-users (farmers, agronomists, nutritionists) to monitor quality metrics, receive real-time alerts, and fine-tune decisions based on model outputs.
- **Advanced Domain Adaptation:** Domain adaptation techniques (e.g., adversarial training, contrastive alignment, style transfer) could enhance model robustness to unseen environments or sensor shifts—critical for low-cost or widely deployed systems.

- **Data-Centric Approaches and Label Efficiency:** Leveraging self-supervised learning and active learning strategies can reduce dependency on expensive labeled data, especially for spectral calibration or segmentation-intensive tasks in novel crops or regions.

Appendix A

Appendix

A.1 Explained Variance for Handheld NIR

Table A.1: Explained variance (%) by each latent variable (LV) for different instruments and target variables. The values represent the individual contribution of each LV to the total explained variance.

Instrument	Variable	LV1	LV2	LV3	LV4	LV5	LV6	LV7	LV8	LV9	LV10
AgroCares Static	ADF	69.64	9.99	1.89	5.29	2.66	2.74	1.73	2.22	0.50	0.80
	ADL	75.36	3.10	3.93	1.58	4.79	2.72	1.73	3.06	0.49	0.36
	CP	34.57	40.04	3.19	4.97	5.35	2.54	2.17	1.89	0.99	0.91
	IVTD	66.68	12.45	1.59	4.12	3.94	3.21	1.77	1.49	1.28	0.83
	NDFD	75.40	4.20	3.90	1.83	2.45	2.76	2.93	2.62	0.80	0.68
	aNDF	38.45	35.27	3.75	5.49	4.13	3.71	2.43	2.03	0.56	0.75
AgroCares Moving	ADF	72.73	11.85	1.76	5.23	2.02	1.75	0.52	1.52	0.26	0.26
	ADL	79.69	3.66	4.60	1.87	2.07	3.65	0.94	1.01	0.17	0.29
	CP	37.47	42.25	2.13	7.73	1.68	3.31	1.06	1.09	0.51	0.21
	IVTD	67.76	15.55	1.78	6.10	1.02	1.57	1.75	1.66	0.29	0.25
	NDFD	78.51	2.07	4.56	2.58	1.23	6.11	1.23	1.18	0.18	0.33
	aNDF	42.04	38.25	2.75	5.86	3.70	1.75	0.70	1.75	0.49	0.21
NEOSpectra Static	ADF	67.24	13.98	7.18	3.96	2.41	0.70	1.23	0.42	0.53	0.32
	ADL	74.43	5.09	6.78	5.11	3.41	0.64	1.10	0.63	0.44	0.31
	CP	36.26	42.49	6.67	5.75	1.79	1.20	0.63	1.28	1.00	0.45
	IVTD	62.96	17.68	5.75	5.34	1.61	1.39	1.68	0.50	0.59	0.34
	NDFD	73.31	2.32	9.77	6.12	1.84	1.48	1.91	0.47	0.43	0.35
	aNDF	36.99	38.83	10.39	4.16	3.23	0.52	1.70	0.45	0.90	0.40
NEOSpectra Moving	ADF	67.69	12.72	4.56	7.72	1.77	1.13	1.12	0.46	0.55	0.28
	ADL	73.90	4.19	8.23	5.41	2.66	0.82	0.45	0.84	1.19	0.28
	CP	35.14	39.27	3.32	13.64	2.10	0.80	1.78	0.98	0.46	0.40
	IVTD	60.10	15.80	2.24	13.97	1.72	1.27	1.47	0.54	0.60	0.28
	NDFD	72.91	1.90	7.88	9.07	2.23	0.87	1.05	1.08	0.73	0.25
	aNDF	38.82	34.14	3.75	15.01	2.22	1.30	0.78	0.85	0.33	0.59
NEOSpectra TurnTable	ADF	65.80	16.21	4.45	6.11	2.14	1.01	0.48	0.87	0.46	0.26
	ADL	72.69	6.89	6.00	3.80	5.30	1.02	0.49	0.31	1.06	0.19
	CP	30.11	50.20	6.68	3.97	2.36	1.91	0.48	1.09	0.58	0.32
	IVTD	59.73	20.88	2.94	7.73	2.67	1.44	0.48	1.05	0.46	0.29
	NDFD	75.94	2.90	7.23	2.55	4.24	3.00	0.50	0.75	0.55	0.19
	aNDF	33.29	45.87	8.35	3.29	2.71	1.56	0.31	1.25	0.46	0.41
Trinamix Static	ADF	74.94	10.20	2.82	2.63	1.33	2.96	1.53	0.70	0.27	0.59
	ADL	79.72	5.00	2.72	1.96	1.78	3.62	1.22	1.35	0.65	0.30
	CP	42.12	37.73	5.05	1.99	2.13	2.41	1.72	3.17	0.45	1.02
	IVTD	70.04	13.63	2.41	3.13	1.39	2.30	0.73	3.34	0.57	0.55
	NDFD	80.21	3.67	3.60	1.17	2.53	2.12	2.01	1.96	0.25	0.54
	aNDF	45.70	33.24	5.78	2.67	1.75	4.03	2.68	0.53	0.70	0.81

ADF - Acid Detergent Fiber, ADL - Acid Detergent Lignin, CP - Crude Protein, IVTD - in-vitro Total Digestibility, NDFD - Neutral Detergent Fiber Digestibility, aNDF - actual Neutral Detergent Fiber