

## Article

# Uncrewed Aerial Vehicle (UAV)-Based High-Throughput Phenotyping of Maize Silage Yield and Nutritive Values Using Multi-Sensory Feature Fusion and Multi-Task Learning with Attention Mechanism

Jiahao Fan <sup>1</sup> , Jing Zhou <sup>2</sup> , Natalia de Leon <sup>3</sup>  and Zhou Zhang <sup>1,\*</sup> 

<sup>1</sup> Biological Systems Engineering, University of Wisconsin—Madison, Madison, WI 53706, USA; jiahao.fan@wisc.edu

<sup>2</sup> Crop and Soil Science, Oregon State University, Corvallis, OR 97331, USA; jing.zhou@oregonstate.edu

<sup>3</sup> Plant and Agroecosystem Sciences, University of Wisconsin—Madison, Madison, WI 53706, USA; ndeleongatti@wisc.edu

\* Correspondence: zzhang347@wisc.edu

## Highlights

### What are the main findings?

- Attention-based deep fusion improved the usage of multi-sensor features while keeping them distinguishable; multi-task learning estimated multiple silage traits with outperformance compared to baselines.
- Hyperspectral data contributed the most to the model; LiDAR and RGB added complementary signals; a retrieval-based option can be used when hyperspectral data are not available.

### What are the implications of the main findings?

- Multi-sensor, attention-based fusion is a practical route to UAV-based, non-destructive phenotyping.
- The trait estimates can be used to support breeding selection and field management while reducing reliance on lab measurements.



Academic Editor: Muhammad Moshir Rahman

Received: 8 October 2025

Revised: 3 November 2025

Accepted: 4 November 2025

Published: 6 November 2025

**Citation:** Fan, J.; Zhou, J.; Leon, N.d.; Zhang, Z. Uncrewed Aerial Vehicle (UAV)-Based High-Throughput Phenotyping of Maize Silage Yield and Nutritive Values Using Multi-Sensory Feature Fusion and Multi-Task Learning with Attention Mechanism. *Remote Sens.* **2025**, *17*, 3654. <https://doi.org/10.3390/rs17213654>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## Abstract

Maize (*Zea mays* L.) silage's forage quality significantly impacts dairy animal performance and the profitability of the livestock industry. Recently, using uncrewed aerial vehicles (UAVs) equipped with advanced sensors has become a research frontier in maize high-throughput phenotyping (HTP). However, extensive existing studies only consider a single sensor modality and models developed for estimating forage quality are single-task ones that fail to utilize the relatedness between each quality trait. To fill the research gap, we propose **MUSTA**, a **M**ulti-Sensory feature fusion model that utilizes **M**ulti-Task learning and the **A**ttention mechanism to simultaneously estimate dry matter yield and multiple nutritive values for silage maize breeding hybrids in the field environment. Specifically, we conducted UAV flights over maize breeding sites and extracted multi-temporal optical- and LiDAR-based features from the UAV-deployed hyperspectral, RGB, and LiDAR sensors. Then, we constructed an attention-based feature fusion module, which included an attention convolutional layer and an attention bidirectional long short-term memory layer, to combine the multi-temporal features and discern the patterns within them. Subsequently, we employed multi-head attention mechanism to obtain comprehensive crop information. We trained MUSTA end-to-end and evaluated it on multiple quantitative metrics. Our

results showed that it is capable of practical quality estimation results, as evidenced by the agreement between the estimated quality traits and the ground truth data, with weighted Kendall's tau coefficients ( $\tau_w$ ) of 0.79 for dry matter yield, 0.74 for MILK2006, 0.68 for crude protein (CP), 0.42 for starch, 0.39 for neutral detergent fiber (NDF), and 0.51 for acid detergent fiber (ADF). Additionally, we implemented a retrieval-augmented method that enabled comparable prediction performance, even without certain costly features available. The comparison experiments showed that the proposed approach is effective in estimating maize silage yield and nutritional values, providing a digitized alternative to traditional field-based phenotyping.

**Keywords:** high-throughput phenotyping (HTP); hyperspectral imagery; light detection and ranging (LiDAR); multi-sensor feature fusion; multi-task learning; attention mechanism

## 1. Introduction

Silage maize (*Zea mays* L.) is extensively cultivated across the globe due to its stable yield performance under diverse environmental and agronomic conditions [1]. As a leading producer, the United States generated 130.32 million tons of maize silage in 2021. Characterized by high energy content and outstanding nutritional properties, maize silage has become the primary ingredient in the feeding regimen of dairy cows for milk production, providing substantial support to the dairy industry [2]. However, maize silage production faces considerable challenges, including weather stress and farmland diminishing [3]. To maintain and stabilize the dairy supply while easing land competition, it is essential to leverage genetic and management innovations that boost maize silage production and nutritive value [4].

Comprising the entire plant, including stalks and digestible grains, maize silage offers high dry matter content and consistent quality [5]. Enhancing silage varieties entails boosting both their production capability and nutritive quality, with the latter reflected in the enzymatic digestibility of the biomass [6]. Superior silage maize varieties are distinguished by their high protein and dry matter content, coupled with high intake potential, which is often associated with a low fiber content [6]. As a result, the nutritive quality of maize silage mainly depends on the dry matter (DM) yield and compositional characteristics, such as crude protein (CP), starch, neutral detergent fiber (NDF), and acid detergent fiber (ADF) concentrations. Typically, breeders integrate these nutritional traits of hybrids using summative criteria, such as the MILK2006 (milk yield per acre index), to determine their potential for conversion into animal productivity units. High-performing hybrids are then selected in breeding cycles to enhance maize germplasm and promote future feed quality [7,8]. However, phenotyping silage nutritive value is still a bottleneck in breeding cycles [9]. Traditionally, these traits are measured by laboratory chemical methods, which are destructive, costly, time-consuming, and may produce hazardous waste [10–12]. In recent years, near-infrared reflectance spectroscopy (NIRS) has been used as a rapid alternative for quantifying forage biochemical traits [13,14]. However, NIRS is still sample-based and labor-intensive, as it requires collecting, drying, and grinding forage material and measuring it on bench-top instruments [15]. NIRS is effective because homogeneous samples show diagnostic absorptions of O–H, C–H, and N–H bonds across roughly 400–2500 nm, which can be calibrated to CP, DM, and fiber fractions [16]. Therefore, in this study, we applied the similar principle at canopy scale using a UAV-borne sensor operating in the visible–NIR window (400–1000 nm). In 400–700 nm, sensor reflectance is mainly governed by pigments and N status, providing indirect sensitivity to protein and maturity;

in 700–1000 nm, reflectance is dominated by canopy structure and greenness, which is informative for fresh/dry matter and biomass [17]. High-throughput phenotyping (HTP) holds the promise of transforming forage phenotyping capacity by supplying researchers with remarkably more phenotypic information than conventional observations. In breeding programs, genetic selection decisions are often made on multiple traits. This creates a practical need for models that can predict multiple related traits from the same UAV data acquisition. Therefore, fusing multiple sensor information and incorporating multi-task learning for simultaneous selection of multiple traits is vital.

Among various sensor options, the optical hyperspectral imager can capture plant interactions with the electromagnetic spectrum across hundreds of narrow-band wavelengths, making it ideal for determining forage biochemical composition. For example, ref. [18] used a UAV-based hyperspectral imager to capture canopy reflectance of grass–legume mixture on grassland to estimate forage fresh (FM) and dry matter (DM) yield, dry matter digestibility (DMD), CP, NDF, and indigestible NDF (iNDF) content with a good estimation performance. Other examples showed that the full wavelength spectrum and narrow-band vegetation indices (VIs) from the UAV-based hyperspectral imagery can be effectively used to estimate the alfalfa quality traits, including forage yield, CP, ADF, and ash-corrected NDF (aNDF) [19,20]. Hyperspectral imaging offers the potential to evaluate forage traits rapidly and non-destructively over large areas; however, the expenses associated with gathering and managing data from hyperspectral imaging are considerable, which curtails its utilization.

In comparison to hyperspectral imagers, optical RGB cameras present a more economically viable method for monitoring crops. Crop canopies can exhibit a range of colors and morphologies due to chlorophyll fluctuations and varying growth conditions, which are associated with crop yield and nutritional values [21–23]. For example, from RGB imagery, by extracting visible indices like the green-red vegetation index (GRVI) and excess green vegetation index (EGVI), leaf nitrogen concentration and above-ground maize biomass can be accurately estimated [24–26]. Additionally, canopy textural and morphological features derived from RGB imagery can reflect changes in vegetation structure and have demonstrated strong correlations with maize leaf area index (LAI) [27], leaf moisture content [28], canopy chlorophyll concentration [29], and plant population [30]. Although successful, most existing studies focus exclusively on single optical data modalities, which may lead to incomplete interpretations of crop productivity and quality. For instance, using single optical sensors might struggle to accurately identify alterations in dense vegetation, owing to asymptotic saturation effects that limit the sensor's ability to respond proportionally during the later stages of crop development [31]. Furthermore, as UAV-based optical imagers capture electromagnetic energy primarily reflected or absorbed from the uppermost layers of crop canopies, discerning differences between crop hybrids that display similar spectral properties yet have distinct canopy heights can be challenging. These limitations highlight the need for additional remote sensing technologies that could serve as a complement to optical imaging, thus providing a more thorough understanding of crop variations and characteristics.

The fusion of LiDAR (light detection and ranging) and optical data offers a solution to the above issues by combining rich spectral and structural information from diverse sensor systems. LiDAR, an active remote sensing technology, can penetrate through different canopy layers, enabling the measurement of both upper and lower canopy features. Mounted on UAVs, LiDAR sensors can provide detailed structural information on crop canopy closure patterns [32] and canopy heights [33], both of which are closely linked to above-ground biomass [34]. In addition to structure-related factors, LiDAR also measures point intensity, reflecting the composition and surface texture of target objects through the

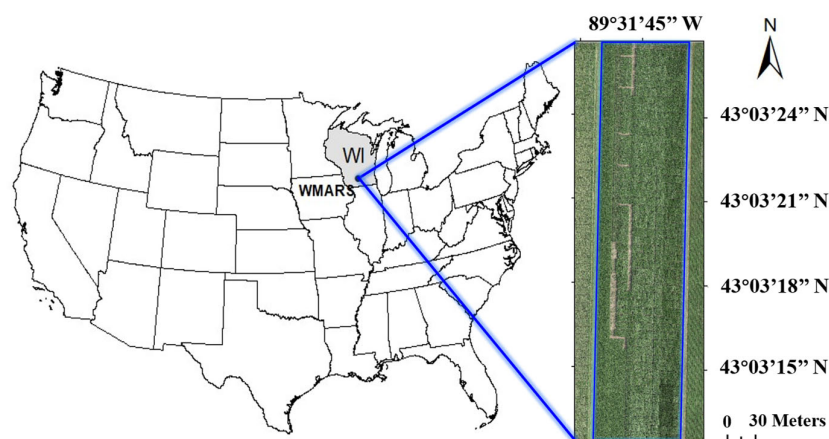
laser return strength [35]. LiDAR–optical data fusion has been widely used in crop HTP. For example, ref. [36] utilized geometric and spectral characteristics from various UAV-based sensors, including an RGB camera, a LiDAR unit, and a push-broom hyperspectral scanner, in support vector regression (SVR) models to predict the end-of-season biomass of sorghum hybrids. In [37], VIs and LiDAR-derived metrics were extracted and partial least squares (PLS) regression was used to estimate maize biomass. However, most feature fusion methods simply concatenate several types of features together, which may not be equally represented or measured, leading to suboptimal model performance, as the model may be overly reliant on the dominant features and fail to capture the information presented by the less represented features. In the context of time-series multi-sensor data, another concern is the increased dimensionality of the stacked features, coupled with the limited availability of labeled samples in many real-world applications, which can lead to the curse of dimensionality and a high risk of overfitting empirical models.

To tackle the identified research gap, we proposed MUSTA, a **M**ulti-Sensory feature fusion model that utilizes **M**ulti-Task learning and **A**ttention mechanisms to simultaneously estimate multiple nutritive values for silage maize hybrids. The main contributions of this study include: (1) we proposed a UAV-based multi-sensor phenotyping workflow that jointly exploits hyperspectral, RGB, and LiDAR data to address the limits of single-modality approaches; (2) we developed an attention-based feature fusion and multi-task learning framework that estimates multiple silage quality traits simultaneously and reveals the relative contribution of each sensor source; and (3) we proposed a retrieval-augmented, low-cost variant method that allows trait estimation when hyperspectral data are missing, improving the practicality of deployment in breeding and field management settings. The findings of this study hold the potential for shaping agricultural management practices and guiding future crop breeding programs.

## 2. Materials

### 2.1. Study Region and Phenotypic Data Collection

The research trials were conducted at the University of Wisconsin's silage maize breeding experiment site within the West Madison Agricultural Research Station (WMARS, Madison, WI, USA. latitude: 43°03'50"N, longitude: 89°32'25"W, Figure 1). WMARS, which served as one of the experimental sites in 2021, typically experiences warm, humid summers and cold, dry winters. A total of 507 silage maize hybrids, each with two field replicates, were cultivated under a randomized complete block design (RCBD). The hybrids, planted in May 2021, were arranged in two-row plots, each 6.5 m long with 0.19 m plant spacing and 0.75 m between rows, and were harvested in September 2021.



**Figure 1.** Silage maize breeding experimental trials at WMARS.

The investigated silage maize varieties were derived from maize inbreds of the UW maize breeding program, including resources from the Germplasm Enhancement of Maize (GEM) project (<https://cornbreeding.wisc.edu/collaborations/united-states-department-of-agriculture-germplasm-enhancement-of-maize-project-usda-gem/>, accessed on 1 September 2025). These included experimental hybrids from both the non-stiff stalk heterotic group and the stiff stalk heterotic group, contributing to the wide genetic diversity in the pool. Established and well-characterized maize lines, such as inbred lines *LH244* and *LH287*, were employed as tester lines. Moreover, a selection of both commercial and internal hybrids was utilized as experimental checks. Phenotypic data were collected at the plot level. Following the final harvest, each plot sample was weighed and taken for moisture and quality assessments. Samples were oven-dried at 60 degrees Celsius for approximately seven days, then they were ground using a Wiley mill to pass through a 1 mm screen. The silage yield was calculated based on a 100% dry matter (DM) basis. The dried samples were then analyzed using a near-infrared spectroscopy, which measures the light absorption at various wavelengths in the near-infrared region. The acquired spectra represent the molecular vibrations of different chemical constituents within the sample [38], including crude protein (CP), starch, neutral detergent fiber (NDF), and acid detergent fiber (ADF) concentrations. The ultimate breeding selection criterion was determined by MILK2006, a milk yield per acre index, calculated by integrating silage DM yield and chemical constituents using a summative equation [39]. A summarized description of each trait can be found in Table 1.

**Table 1.** Descriptions of maize phenotypic traits and their measurements.

| Phenotypic Traits | Unit        | Measuring Description                                                                                                                                                                         |
|-------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DM Yield          | US ton/acre | Measured by weighing a sample in the field, putting it in a dryer, weighing it again after it has dried to a constant weight, and calculating DM yield using DM percentage at the plot level. |
| CP                | %           | Measured by NIRS, expressed as a percentage of DM content.                                                                                                                                    |
| Starch            | %           | Measured by NIRS, expressed as a percentage of DM content.                                                                                                                                    |
| NDF               | %           | Measured by NIRS, expressed as a percentage of DM content.                                                                                                                                    |
| ADF               | %           | Measured by NIRS, expressed as a percentage of DM content.                                                                                                                                    |
| MILK2006          | US ton/acre | Milk yield per acre index, calculated by the NIRS-obtained quality values.                                                                                                                    |

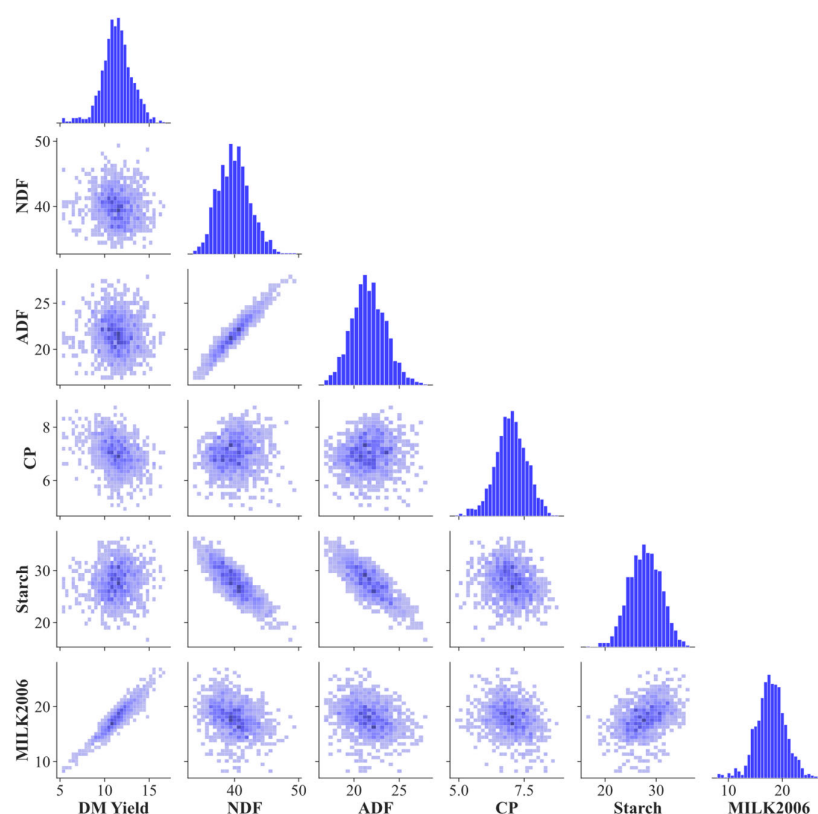
## 2.2. Phenotypic Data Exploration

Six quality traits, including DM yield, CP, starch, NDF, ADF, and MILK2006, were considered in this study. Table 2 summarizes the statistics of maize silage phenotype harvested at the end of the season, showing low coefficients of variation (<20%) for all six phenotypes, indicating low data dispersion relative to the means. For the classification task, the 33rd and 66th percentiles are used as thresholds to determine the levels of quality values. Figure 2 displays the distribution histograms of each trait (in diagonal) and correlation degrees between each pair of traits (in scatter plots). Each of the traits basically follows a normal distribution. Moreover, it shows that the pairs of DM yield–MILK2006 and NDF–ADF exhibit a positive linear relationship, suggesting that, for instance, a higher NDF concentration often accompanies a higher ADF concentration. Conversely, the pairs ADF–starch and NDF–starch present a negative linear relationship, suggesting that, for example, a higher NDF concentration typically corresponds to a lower starch concentration. These relationships inspire us to develop multi-task models that can share parameters between related prediction tasks.

**Table 2.** Descriptive statistics for each observed maize silage phenotype, gathered at the end of the growth season.

| Phenotypic Traits         | Count | Mean  | Std * | CV ** | Min   | 33%   | 50%   | 66%   | Max   |
|---------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| DM Yield<br>(US ton/acre) | 1240  | 11.37 | 1.67  | 14.70 | 5.22  | 10.73 | 11.37 | 11.96 | 16.80 |
| CP (%)                    | 1240  | 6.99  | 0.60  | 8.59  | 4.86  | 6.76  | 7.00  | 7.23  | 8.81  |
| Starch (%)                | 1240  | 27.78 | 3.08  | 11.10 | 16.31 | 26.32 | 27.76 | 29.12 | 36.52 |
| NDF (%)                   | 1240  | 39.95 | 2.49  | 6.24  | 33.52 | 38.78 | 39.85 | 40.91 | 49.64 |
| ADF (%)                   | 1240  | 21.63 | 1.90  | 8.77  | 16.70 | 20.80 | 21.57 | 22.35 | 28.11 |
| MILK2006<br>(US ton/acre) | 1240  | 17.88 | 2.76  | 15.43 | 7.95  | 16.81 | 17.86 | 18.97 | 20.09 |

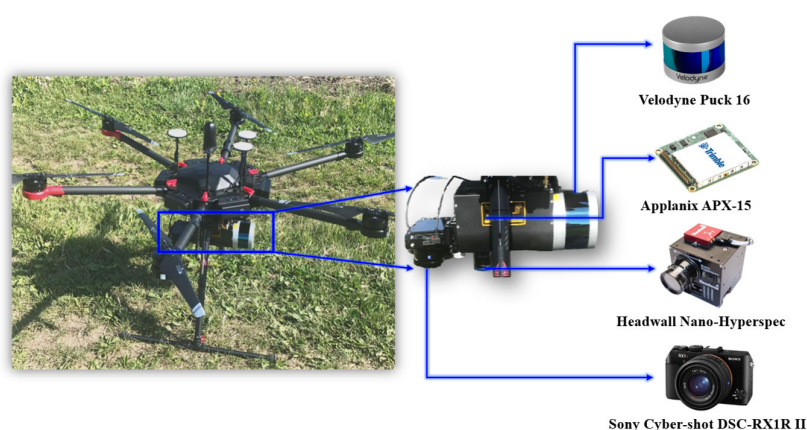
\* Standard deviation, \*\* coefficient of variation.

**Figure 2.** Distribution histograms of maize silage quality traits (in diagonal) and scatter plots between pairs of traits.

### 2.3. Multi-Sensory Data Collection and Pre-Processing

In this study, a UAV-based imaging system was developed using the DJI M600 Pro hexacopter (DJI Technology Co., Shenzhen, China) as the platform (Figure 3). The UAV platform carried three image sensors, including a Headwall Nano-Hyperspec push-broom hyperspectral scanner (Headwall Photonics Inc., Bolton, MA, USA), a Velodyne Puck 16 LiDAR unit (Phoenix LiDAR Systems, Los Angeles, CA, USA), and a Sony Cyber-shot DSC-RX1R II digital RGB camera (Sony Corporation, Tokyo, Japan). Furthermore, the APX-15 (Applanix Corporation, Richmond Hill, OR, Canada) global navigation satellite system and inertial measurement unit (GNSS/IMU) were utilized for data georeferencing. The onboard GNSS/IMU system is essential for the effective co-registration of multi-modal data. The GNSS/IMU system, integrating both a Global Navigation Satellite System and an Inertial Measurement Unit, was processed to determine exact UAV positions and orientations for each data collection timestamp. These coordinates then were used to georeference and synchronize data from different sensors, including LiDAR, RGB cameras, and hyper-

spectral scanners. To refine trajectory data further, we employed post-processing software named POSPac (v 8.6.7810.21805, Applanix Corporation, Richmond Hill, OR, Canada), which utilized Post-Processing Kinematic (PPK) techniques and real-time corrections from continuously operating reference stations (CORSS) to enhance GNSS/IMU data accuracy. Additionally, the high accuracy of the GNSS system facilitated the creation of Digital Surface Models (DSMs) from LiDAR data. DSMs afterwards enabled precise mapping of image points onto a physical coordinate system, ensuring feature alignment across the multimodal dataset. The description of the sensors used in this study is provided in Table 3. Meticulous calibrations were carried out to ensure the spatial and temporal multi-sensor co-alignment [36,40]. A pre-defined flight plan was executed using a flight control app named DJI GS Pro (DJI Technology Co., Shenzhen, China), flying at an altitude of 60 m and a speed of 6 m/s, with a 12 m distance between each flight path. Seven UAV surveys were conducted to acquire multi-sensory data between June and September 2021 under clear and calm weather conditions, encompassing important maize vegetative and reproductive stages. The survey dates and corresponding days after sowing (DAS) are presented in Table 4. The subsequent sections provide an in-depth discussion on data preprocessing. It is worth noting that individual plot segmentation is performed by manually drawing plot boundaries (plot-level scale), and then batch-processed using the ExtractByMask function from the ArcPy library in ArcGIS Pro (V3.4.0). Features extracted from each modality were subsequently aggregated based on plots. Therefore, feature-level co-registration was ultimately harmonized in feature space at the plot level.



**Figure 3.** The UAV imaging platform and onboard sensors.

**Table 3.** The imaging sensors description.

| Sensor                | Description                                                                                                                                                     |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Hyperspectral scanner | Headwall Nano-Hyperspec, 274 spectral channels ranging from 400 to 1000 nm, 2.2 nm bandwidth, 12.7 mm lens, $640 \times 1$ pixels, $7.4 \mu\text{m}$ pixel size |
| LiDAR unit            | Velodyne Puck 16, 16 channels, 600 rotations per minute (RPM), 100 m maximum range                                                                              |
| RGB camera            | Sony Cyber-shot DSC-RX1R II, 3 channels, 35 mm lens, $7952 \times 5304$ pixels, $4.52 \mu\text{m}$ pixel size                                                   |

**Table 4.** The UAV survey dates and days after sowing (DAS).

| Survey Date (dd/mm/yyyy) | DAS |
|--------------------------|-----|
| 27 June 2021             | 53  |
| 2 July 2021              | 58  |
| 17 July 2021             | 73  |
| 30 July 2021             | 86  |
| 13 August 2021           | 100 |
| 21 August 2021           | 108 |
| 30 August 2021           | 117 |

### 2.3.1. Hyperspectral Imagery

Hyperspectral data were obtained using a Headwall nano-hyperspec push-broom scanner (Headwall Photonics, Inc., Bolton, MA, USA), which featured 274 spectral bands ranging from 400 to 1000 nm and a bandwidth of 2.2 nm. The flight missions were conducted under clear and calm weather conditions, with a 46% lateral overlap between flight passes for hyperspectral imaging, yielding a 3.50 cm ground sampling distance (GSD). After data acquisition, two primary pre-processing steps were performed: geometric correction and radiometric correction. Geometric correction was achieved by orthorectifying the hyperspectral data using the GNSS/IMU data in GRYFN software (v1.6.7, GRYFN, Inc., West Lafayette, IN, USA). For radiometric correction, raw digital numbers (DNs) were converted into radiance using SpectralView software (v3.3.0.1, Headwall Photonics, Inc., Bolton, MA, USA), followed by radiometric calibration using reference panels with reflectance of 56%, 32%, and 11%. Image backgrounds (e.g., shadows and soil) were removed by setting a threshold in the NIR band, as vegetation typically exhibits higher reflectance in the NIR region compared to the background. In this research, hyperspectral pixels with reflectance below a 15% threshold at an 800 nm wavelength were removed.

### 2.3.2. LiDAR Data

UAV-based LiDAR data were collected using a Velodyne LiDAR PUCK-16 system (Phoenix LiDAR Systems, Los Angeles, CA, USA), emitting pulses at a wavelength of 903 nm and a frequency of 5 kHz. The flight missions were conducted under clear and calm weather conditions, with an 83% lateral overlap between flight lines for LiDAR sensing, resulting in an approximate single pass point density of 116 points/m<sup>2</sup>. The LiDAR system, equipped with a GNSS/IMU unit, recorded the point positions in real time and generated point clouds with a geometric accuracy of  $\pm 3$  cm. The cloth simulation filtering method (CSF) was employed to distinguish ground and plant points after obtaining the point clouds [41]. By employing separated ground points, digital terrain models (DTMs) with an 8 cm resolution were created using the bilinear interpolation method. The DTM represents the terrain elevation, which is assumed to be consistent throughout the growing season. Lastly, relative heights of plant canopy points were calculated to generate crop height models (CHMs) by subtracting the DTM from the altitude values of plant canopy points. Noisy points outside the range of 0–4 m were removed.

### 2.3.3. RGB Imagery

UAV-based ultra-high-resolution RGB imagery was obtained using a Sony Cyber-shot DSC-RX1R II digital camera (Sony Corporation, Tokyo, Japan). The UAV ultra-high-resolution RGB data were captured under stable lighting conditions. The flight missions were conducted under clear and calm weather conditions, with a 78% forward overlap and an 81% lateral overlap between flight lines, yielding RGB imagery with a 0.77 cm GSD. The images featured three bands: red (*r*), green (*g*), and blue (*b*). To minimize the impact of lighting and shading, the RGB imagery's color space was normalized. Normalization can remove highlights and shadows, facilitating the comparison of image analysis results obtained at different times. The normalization method was applied using Equation (1), where *r*, *g*, and *b* represent the original DN values. The original DN values were converted into normalized values ranging from 0 to 1. To extract only vegetation pixels, image backgrounds (e.g., shadows and soil) were removed using the excess green (ExG) index [42]. ExG was calculated using Equation (2), and an ExG value greater than 0 indicated vegetation pixels.

$$R = r / (r + g + b), G = g / (r + g + b), B = b / (r + g + b) \quad (1)$$

$$\text{ExG} = 2G - R - B \quad (2)$$

### 3. Methodology

In this study, UAV-based hyperspectral, RGB, and LiDAR data were used to estimate DM yield and various nutritive values for silage maize hybrids. Figure 4 provides a visual representation of the overall workflow of this study. The estimation procedure consisted of six steps: (1) data collection; (2) data pre-processing; (3) feature extraction; (4) model development and evaluation; (5) permutation feature importance analysis; (6) retrieval-augmented quality estimation. We discuss the UAV-based data and ground truth data acquisition and subsequent data pre-processing in Section 2.2. The following sections will describe feature extraction, model development, feature importance analysis, and estimation accuracy assessment.

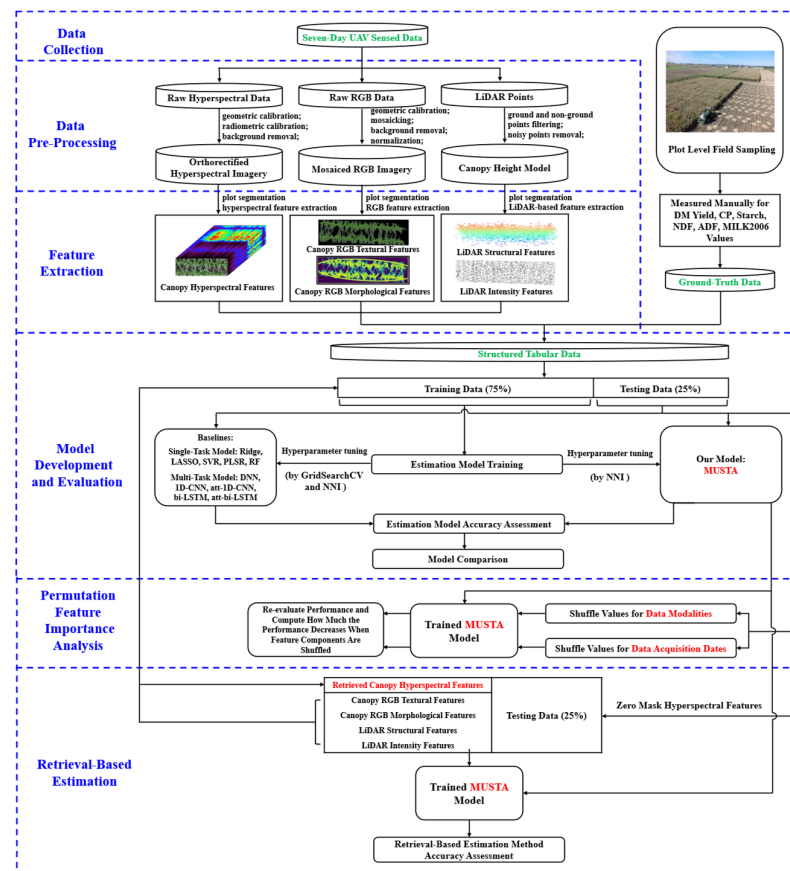


Figure 4. The overall workflow of this study.

#### 3.1. Feature Extraction

In this study, we focused on both optical-based and LiDAR-based time-series features for their relevance in estimating maize silage quality traits, including canopy hyperspectral features ( $n_h = 274$ ), canopy RGB textural features ( $n_t = 20$ ), canopy RGB morphological features ( $n_m = 25$ ), LiDAR structural features ( $n_s = 36$ ), and LiDAR intensity features ( $n_i = 16$ ), spanning seven survey days ( $t_s = 7$ ). Consequently, each processed sample plot yielded a total of  $N = (n_h + n_t + n_m + n_s + n_i) \times t_s = 2597$  features. Interpreting each sample plot as an individual instance (row) and each extracted feature as an attribute (column), the curated dataset can be regarded as structured tabular data. Further details on the extracted features are provided in the succeeding sections.

### 3.1.1. Hyperspectral-Based Features

**Canopy Hyperspectral Features:** Hyperspectral reflectance signifies the absorption of light at particular wavelengths, which is connected to plant properties. As demonstrated in [31], the phenotypic traits of numerous hybrids might be too subtle to differentiate without collectively considering the full-band reflectance spectra. Therefore, in this research, we extracted the average reflectance values at various bands from the hyperspectral imagery and smoothed the full band spectral profile using the Savitzky–Golay filter. We selected a window length of 9 and a polynomial degree of 3 for the Savitzky–Golay filter to effectively reduce noise while preserving key spectral features. These parameters were fine-tuned by minimizing the root mean square error of predictions using a single-source data model, specifically utilizing hyperspectral data to predict the MILK2006. This resulted in a total of 274 reflectance values for each plot per UAV survey.

### 3.1.2. RGB-Based Features

**Canopy RGB Textural Features:** Maize canopy optical textural features depict the spatial organization, patterns, and variations among the RGB image pixels of the maize plant canopy. They are widely used for crop lodging prediction, crop classification, and crop growth status evaluation [43–46]. Textural information is commonly expressed using properties calculated from the Gray Level Co-occurrence Matrix (GLCM), which describe the spatial relationships between pairs of pixels with specific grey levels, separated by a certain distance  $d$  and angle  $\theta$ . The GLCM can be computed using the `greycomatrix` function built in the `skimage` Python Library (v3.10.19) [47], while the properties containing the image’s textural features can be computed using the `greycoprops` function. In calculating the GLCM, we set the distance  $d$  to 1 and selected four directions for  $\theta$ :  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ . From these matrices, we extracted textural features of contrast, correlation, energy, homogeneity, and dissimilarity. Consequently, we obtained a total of 20 textural features for each plot per UAV survey.

**Canopy RGB Morphological Features:** Maize canopy morphological features refer to the shape and distribution of the maize plant canopy, which can be analyzed to gain insights into plant growth, development, and overall health. The shape and distribution of the maize canopy are indicative of the plant’s competition for resources and its capacity to intercept and absorb sunlight, essential for photosynthesis and energy production, which in turn may influence forage quality [30,48,49]. While many studies have focused on the morphology of individual plants (like leaf size and stem structure), few have investigated the impact of morphological features on forage yield and quality at the plot level. In this study, we refined and adapted the shape feature extraction module presented in [50] to extract 25 distinct crop canopy morphological features, including solidity, rectangularity, breadth, circularity, roundness, eccentricity, sphericity, etc., for each segmented plot during each UAV survey. A more detailed RGB-based feature list is shown in Appendix A.

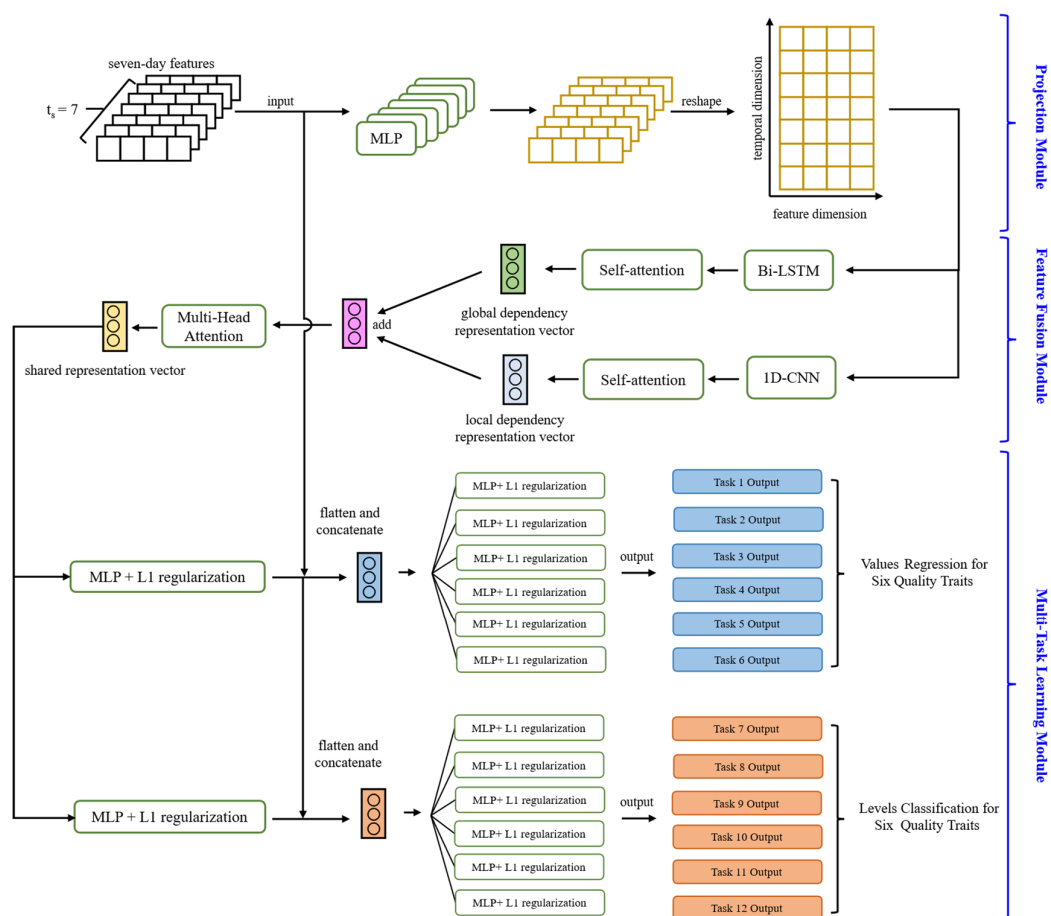
### 3.1.3. LiDAR-Based Features

**LiDAR Structural Features:** Maize structural features represent the three-dimensional attributes of maize plants, encompassing canopy volume, surface roughness, and layering, which can reveal information about the maize plot’s overall architecture. Utilizing the CHM model outlined in Section 2.3.2, we adhered to the process presented in [51–54] and extracted various maize structural features, including crop height percentiles, canopy cover, crop height statistics, canopy volume, projected leaf area (PLA), plant area index (PAI), and plant area density (PAD). As a result, we acquired a total of 36 structural features for each plot during each UAV survey.

**LiDAR Intensity Features:** LiDAR-based intensity features depict the intensity of the returned LiDAR laser pulses, which reflect the target surface's properties. Specifically, they provide information about characteristics like reflectance intensity, which is the ability of the surface to reflect laser light back to the sensor, and texture, which describes the physical composition and variation in the surface. These intensity features offer supplementary information about the maize canopy, complementing the structural features derived from LiDAR data [35]. Using the LiDAR points, we extracted LiDAR point number, LiDAR intensity percentiles, and intensity statistics. As a result, we gathered a total of 16 intensity features for each plot during each UAV survey. A more detailed LiDAR-based feature list is shown in Appendix B.

### 3.2. Proposed Model

MUSTA is composed of three main modules: (1) a feature projection module, (2) a feature fusion module, and (3) a multi-task learning module. The model architecture is illustrated in Figure 5. The multi-temporal features, extracted from the seven-day UAV survey, were utilized as inputs for the input layer. Then, these seven-day features were processed through a projection module containing seven multilayer perceptrons (MLPs) to learn and capture different patterns within each single day. Next, these adjusted inputs were fed into the feature fusion module, which employed self-attention and a multi-head attention mechanism to focus on the most relevant features and learn a shared fusion feature set. This fused feature set was then fed into the multi-task learning module, which enabled quality values regression and quality levels classification for six quality traits. The MUSTA model was implemented using Keras in Python.



**Figure 5.** The architecture of MUSTA.

**Projection Module:** This module was designed to identify and encapsulate various patterns in the feature set of each individual day. Initially, the multi-sensory features on each day were regarded as separate vectors and were input into seven distinct MLPs. The outputs from these MLPs were then concatenated and reshaped into a format that was compatible with the 1D-CNN and bi-LSTM layers present in the subsequent feature fusion module.

**Feature Fusion Module:** This module was designed to effectively combine multi-temporal features and learn the underlying coherence patterns within them. It comprises an attention-based 1D-convolutional (att-1D-CNN) layer, an attention-based bidirectional long short-term memory (att-bi-LSTM) layer, and a multi-head attention (MHA) layer. The 1D-CNN layer extracts local dependencies from sequential data by applying convolutional kernels, which are small sliding windows that move across the input data. Local dependencies are those relationships or patterns that occur within the defined time windows of the data. Meanwhile, bi-LSTM captures global dependencies by recognizing relationships or patterns that might occur across wider spans of the input data. bi-LSTM is an enhanced version of the traditional LSTM architecture, combining two LSTM networks that operate in opposite directions—one from the past to the future and the other from the future to the past. This signifies that a prediction at a given timestep relies on the context from steps both preceding and succeeding it. Bi-LSTM excels at capturing these dependencies, given that its architecture, inclusive of a memory cell and a forget gate, allows it to forget, retain, and update these long-term dependencies.

In the context of maize plants, their growth status may undergo swift changes within a brief period, influenced by external environmental factors and their inherent genetic characteristics [55]. Hence, 1D-CNN essentially empowers the network to detect such nearby patterns and short-term dependencies within the temporal multi-sensory features. As for long-term dependencies, the end-of-season yield and quality values depend not only on the growing conditions within a specific brief timeframe, but also on conditions from weeks or even months preceding or succeeding that timeframe. Acknowledging these long dependencies captured by bi-LSTM is also crucial for enhancing prediction accuracy.

Moreover, we topped a self-attention mechanism [56] on these local dependencies and global dependencies information extraction layers to prioritize relevant information extracted and focus on the most relevant ones. It does this by creating a weighted combination of all timesteps in the time-series data, where the weights indicate the relevance of other timesteps to the current timestep. Additionally, a multi-head attention layer was employed to extract comprehensive crop information from the addition of outputs of the att-1D-CNN and att-bi-LSTM layers. Multi-head attention is an extended form of the self-attention mechanism. Its main concept is to perform a self-attention process multiple times in parallel, with each instance employing various learned linear transformations of the original input vectors, thus focusing on different positions, and capturing various aspects of relationships in the data. We used multi-head attention to establish a shared representation for multiple quality estimation tasks.

**Multi-task Learning Module:** This module was designed to process comprehensive crop information and generate predictions. The shared representation obtained from the feature fusion module was then channeled into two separate branches, each representing a distinct task series: quality values regression and quality levels classification. The proposed model was designed based on hard parameter sharing—each branch comprised a general MLP layer and several task-specific output MLP layers. To avoid model overfitting, we introduced L1-norm regularization after each MLP. Notably, to tackle the potential vanishing gradient problem in training extremely deep neural networks, we utilized the shortcut connection [57] before the task-specific MLPs, which allowed for direct backpropagation

of the gradient to the earlier layers in the network. We accomplished the shortcut connection by concatenating the original input from the input layer with the output of the general MLP layers, and then feeding them into the task-specific layers in the multi-task learning module.

**Loss Functions:** In multi-task learning, a unified model is constructed to handle multiple tasks simultaneously. This process involves the use of a combined loss function that is generally an aggregate—often a sum or weighted sum—of the individual loss functions of each task. In this study, we adopted the sum loss within both task branches. The formulas of the sum loss for regression and classification tasks are shown in Equations (3) and (4):

$$J(w_{reg}) = \frac{1}{N} \times \sum_t^T \sum_n^N (y_n - \hat{y}_n)^2 \quad (3)$$

$$J(w_{cls}) = -\frac{1}{N} \sum_t^T \sum_n^N \sum_q^Q c_n \times \log(\hat{c}_n) \quad (4)$$

where  $J$  is designed as a sum combination of the task-specific cost function;  $w_{reg}$  and  $w_{cls}$  are the model parameters for the regression task branch and the classification task branch, respectively;  $T$  is the number of tasks in the task branch;  $N$  is the number of samples;  $y_n$  and  $\hat{y}_n$  represent the observed and the predicted value of sample  $n$ ;  $Q$  is the number of categories in the classification task branch;  $C_n$  and  $\hat{C}_n$  represent the observed and the predicted category of sample  $n$ .

**Hyper-parameter Tuning:** We use the Microsoft NNI (Neural Network Intelligence, [nni.readthedocs.io](https://nni.readthedocs.io), accessed on 1 September 2025) toolkit to automatically tune the model hyper-parameters. The NNI toolkit enabled automated hyper-parameter tuning and efficient hyper-parameter searching for our deep learning model. We utilized the built-in Tree-structured Parzen Estimator (TPE) tuner, a Bayesian optimization approach known for its effectiveness in finding optimal hyper-parameter values. TPE identifies optimal hyper-parameter values by building a probabilistic model of the objective function, enhancing the search for the best configuration. We selected the hyper-parameters that exhibited the best performance for MUSTA.

### 3.3. Comparative Evaluation and Performance Metrics

In this study, five traditional single-task machine learning models and five deep learning-based multi-task models were compared as baselines in estimating the maize silage quality values.

**Baselines:** We compared MUSTA with the following traditional single-task machine learning models: ridge regression (Ridge), least absolute shrinkage and selection operator regression (LASSO), support vector regression (SVR), partial least squares regression (PLSR), random forest regression (RF); and the following deep learning-based multi-task models: deep neural network (DNN), 1-D convolutional neural network (1D-CNN), attention-based 1D-convolutional neural network (att-1D-CNN), bidirectional long short-term memory neural network (bi-LSTM), attention-based bidirectional long short-term memory neural network (att-bi-LSTM). The traditional machine learning-based single-task models were developed using Scikit-learn [58] in Python, and the deep learning-based multi-task models were developed using Keras in Python.

**Hyper-parameter Tuning:** The hyper-parameters in single-task traditional machine learning models were optimized using the GridSearchCV method in Scikit-learn. For Ridge and LASSO regression, we explored regularization strengths ranging from 0 to 10 with a step size of 0.01. In Support Vector Regression (SVR) with an RBF kernel, we tested penalty values from 0.0001 to 11, with a step size of 0.01. For Partial Least Squares Regression (PLSR), we evaluated the number of components from 1 to the feature length. Lastly, in Random

Forest, we experimented with the number of decision trees between 500 and 1000. These parameter ranges were searched based on cross-validation to ensure fair comparison. The hyper-parameters for the deep learning-based multi-task baseline models were optimized using the *NNI* toolkit, the same as the approach used for the MUSTA model.

**Performance Evaluation:** We employed quantitative metrics collectively to ensure a comprehensive evaluation of the model's performance in predicting yield and nutritive values. In typical maize breeding processes, resources are limited, and only a specific proportion of superior breeding materials, such as the top 10%, are selected for further evaluation [59]. Therefore, our primary goal is to preliminarily and accurately identify the rankings and the top-performing subset of hybrids. Motivated by this, the weighted Kendall's tau coefficient ( $\tau_w$ ), Pearson coefficient ( $r$ ), mean absolute error (MAE), and root mean square error (RMSE) metrics were used as metrics for evaluating the regression models' performance. It is worth noting that  $\tau_w$  is a variant of Kendall's tau, which is used for evaluating the correlation between rankings, with more significance attributed to higher-ranked observations via a weighting function. In this study, the weighting function has been calibrated to favor high-performance hybrids displaying quality attributes as high values of DM yield, MILK2006, CP, and starch, alongside low values of NDF and ADF. Consequently, when models are assessed using  $\tau_w$ , it effectively mirrors the model's competence in identifying top-rated hybrids, as evidenced by the agreement in rankings between predicted and actual values for each quality attribute. The formulas of the regression metrics are shown in Equations (5)–(8):

$$\tau_w = \frac{\sum_{n < m} w(n, m) \times \text{sign}(y_n - y_m) \times \text{sign}(\hat{y}_n - \hat{y}_m)}{\sum_{i < j} w_{ij}}$$

where

$$\text{sign}(x) := \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases} \quad (5)$$

$$r = \frac{\sum_n (y_n - \bar{y}) \times (\hat{y}_n - \bar{\hat{y}})}{\sqrt{\sum_n (y_n - \bar{y})^2 \times \sum_n (\hat{y}_n - \bar{\hat{y}})^2}} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \times \sum_n (y_n - \hat{y}_n)^2} \quad (7)$$

$$\text{MAE} = \frac{1}{N} \times \sum_n |y_n - \hat{y}_n| \quad (8)$$

where  $N$  is the number of samples;  $y_n$  and  $\hat{y}_n$  represent the observed and the predicted value of sample  $n$ ;  $y_m$  and  $\hat{y}_m$  represent the observed and the predicted value of sample  $m$ ;  $w(\cdot)$  denotes a weight function which is bounded and symmetric; we adopted the weighting scheme from [60];  $\bar{y}$  and  $\bar{\hat{y}}$  denotes the mean of the observed and the predicted values. Models with higher  $r$  and  $\tau_w$ , and lower RMSE and MAE indicate better performance of a model.

Meanwhile, we evaluated the multi-task model's classification performance using metrics including average accuracy ( $\text{Accuracy}_{avg}$ ), recall score for the best-performing category ( $\text{Recall}_{top\_c}$ ), precision score for the best-performing category ( $\text{Precision}_{top\_c}$ ), and F1 score for the best-performing category ( $\text{F1}_{top\_c}$ ). Average accuracy in multi-class settings calculates the average of the accuracies obtained for every individual category. The recall score quantifies the model's proficiency in identifying all relevant instances of the same category in a dataset. Precision, on the other hand, measures the accuracy of items correctly identified as part of the positive category. F1 score balances the use of precision and

recall, giving a single metric that considers both. As our primary goal is to identify the rankings and the top-performing subset of hybrids, we computed  $Recall_{top\_c}$ ,  $Precision_{top\_c}$  and  $F1_{top\_c}$  to gauge the models' effectiveness in pinpointing these top-performing hybrids. The formulas of the classification metrics are shown in Equations (9)–(13):

$$Accuracy_q = (TP_q + TN_q) / Total\_Observations \quad (9)$$

$$Accuracy_{avg} = \frac{1}{Q} \sum_q Accuracy_q \quad (10)$$

$$Recall_{top\_c} = TP_{top\_c} / (TP_{top\_c} + FN_{top\_c}) \quad (11)$$

$$Precision_{top\_c} = TP_{top\_c} / (TP_{top\_c} + FP_{top\_c}) \quad (12)$$

$$F1_{top\_c} = 2 \times (Precision_{top\_c} \times Recall_{top\_c}) / (Precision_{top\_c} + Recall_{top\_c}) \quad (13)$$

where  $TP_q$  is the number of true positive predictions for category  $q$  made by the model;  $TN_q$  is the number of true negative predictions for all other categories made by the model;  $Total\_Observations$  is the total number of instances;  $Q$  is the number of categories in the classification task branch;  $TP_{top\_c}$  is the number of true positive predictions for the top-performing categories made by the model;  $FN_{top\_c}$  is the number of false negative predictions for the top-performing category made by the model;  $FP_{top\_c}$  is the number of false positive predictions for the top-performing category made by the model.

All models were evaluated using a four-fold cross-validation method, with a random split of 75% of the data as the training dataset and 25% as the test dataset in each fold.

**Class Separability of Fused Features vs. Stacked Features:** To assess the class separability, we employed linear discriminant analysis (LDA) to analyze the fused features and the original stacked features. LDA is a supervised statistical technique extensively utilized in machine learning for dimensionality reduction and classification. The primary objective of LDA is to maximize the distance between the class centroids (between-class variance), while simultaneously minimizing the scatter within each class (within-class variance). We applied LDA by setting the number of components to two, projecting the high-dimensional data into a two-dimensional space for easier visual interpretation of class separability. To assess class separability, we calculated the average position of all data instances within each class cluster to determine the centroid. By computing the Euclidean distance between the centroids of the class clusters formed by LDA, we investigated the extent of class separability between the fused features and stacked features. We aim to determine if the fused features contribute to improved recognition of the best-performing category amidst a multitude of hybrids.

### 3.4. Permutation Feature Importance

Permutation feature importance (PFI) is a technique used to determine the most important features globally in a machine learning or deep learning model by permuting the values of features and scoring the impact on the model performance. It can be applied to a fitted model where the data comes in a tabular form. The concept revolves around assessing the influence of each feature on the model's predictive strength by shuffling the values for those features and measuring how much the performance decreases.

In this study, we first trained our model and measured the  $\tau_w$  coefficient on the test data as the benchmark performance. Then, we randomly shuffled the feature values based on (1) UAV sensor modality and (2) UAV survey timing across the test data. This permutation strategy was inspired by conditional Permutation Importance—instead of permuting a single feature, we permuted feature sets within a day or a data modality to reduce the side effects of spectrally correlated features. We performed five iterations of the

PFI process before calculating the average importance scores, with each iteration starting from a different random seed, aiming to minimize the impact of randomness inherent in the PFI process. Following this, we re-evaluated our model's performance using the rearranged test data and calculated the gap between the benchmark performance and the permuted feature performance ( $(base\_score - permuted\_score) / base\_score \times 100\%$ ). This will typically result in a reduction in performance, and this reduction provides an indication of the importance of each data modality and data acquisition date. Larger discrepancies imply that the feature carries greater weight in the model's decision-making process. The pseudo-code of this PFI analysis is in Algorithm 1.

---

**Algorithm 1:** PFI Analysis Pseudo-code, Python-like

---

```

"""
hs_cols: hyperspectral feature columns
tx_cols: textural feature columns
morp_cols: morphological feature columns
struc_cols: structural feature columns
int_cols: LiDAR intensity feature columns
date_x: columns of features belong to UAV survey date x
"""

for data_modality in [hs_cols, tx_cols, morp_cols, struc_cols, int_cols]:
    get_importances(X, y, data_modality)

for data_acquisition_date in [date_1, date_2, ...]:
    get_importances(X, y, data_acquisition_date)

def get_importances(X, y, columns_to_shuffle):
    """
    columns_to_shuffle: is a sequence of column numbers to shuffle
    """
    base_score = score_func(X, y)
    permuted_X = feature_shuffling(X, columns_to_shuffle)
    permuted_score = score_func(permuted_X, y)
    feature_importance = (base_score - permuted_score) / base_score
    return feature_importance

def score_func(X, y):
    y_pre = MUSTA.predict(X)
    score = compute_weighted_tau(y, y_pre)
    return score

```

---

### 3.5. Retrieval-Augmented Quality Traits Estimation

Hyperspectral imaging, on the other hand, offers the potential to evaluate these traits rapidly and non-destructively over large areas. However, the equipment cost and data processing cost are high, limiting the widespread use of hyperspectral devices [61]. Against this backdrop, we proposed a retrieval-based method, which enables practical estimations even in the absence of costly hyperspectral features, presenting a budget-friendly alternative for breeding or farming missions.

The retrieval-based method hinges on the acquisition of similar hybrid plots from the training database. Firstly, based on existing canopy textural features, canopy morphological

features, LiDAR structural features, and LiDAR intensity features, the model identifies the top 10 analogous plots to a testing sample plot from the training database, using the nearest neighbor search algorithm. Subsequently, the model calculates the cosine similarities between the test sample plot and these ten nearest neighbors, treating these similarity values as the weights. These weights were then used to compute a weighted average of the top 10 analogous plots' hyperspectral features, representing the retrieved hyperspectral features for the testing sample plot. Following this, the model integrates the retrieved hyperspectral features with the already presented features of the sample plot. This combined data was fed into the feature extraction layers and multi-task layer for retrieval-augmented predictive values. The pseudo-code of this retrieval-based method is in Algorithm 2.

---

**Algorithm 2:** Retrieval-Augmented Quality Traits Estimation Pseudo-code, Python-like

---

```
def retrieval_estimation(training_db, test_plot):
    """
    training_db: tabular data, has all the feature columns, including the hyperspectral
    features(hs_col), textural features, morphological features, structural features, LiDAR
    intensity features(none_hs_col).

    test_plot: tabular data, has the same columns as training_db, but the hyperspectral
    features are blank (zero-masked).
    """
    hs_retrived = retrieve_hs_feature(training_db, test_plot)
    test_plot[hs_col] = hs_retrived
    retrieval_estimation_values = MUSTA.predict(test_plot)
    return retrieval_estimation_values

def retrieve_hs_feature(training_db, test_plot):
    analogous_plots_n = nearest_nb_search(training_db[none_hs_col],
    test_plot[none_hs_col], n_nbs=10)
    w_n = cos_similarity(analogous_plots_n[none_hs_col], test_plot[none_hs_col])
    hs_retrived = sum (analogous_plots_n[hs_col] * w_n)/sum (w_n)
    return hs_retrived
```

---

## 4. Results

### 4.1. Model Comparison and Performance

In this section, we assess the regression performance of all models to discern their respective strengths and weaknesses.

#### 4.1.1. Model Regression Performance

The performances of the five single-task traditional machine learning models and six deep learning-based multi-task models are shown in Appendix C, Table A1. Specifically, the results show:

**Single-Task Models:** Ridge regression demonstrates a relatively high performance in predicting DM yield and MILK2006 with  $\tau_w$  values of 0.75 and 0.63, respectively, indicating a strong positive ranking correlation. However, the model's performance decreases significantly for NDF, ADF, and starch. LASSO also shows a similar trend as ridge regression in a performance decrease for NDF but underperforms ridge regression for MILK2006 ( $\tau_w = 0.71$ ). The SVR model's performance is inconsistent, with a marked low of  $\tau_w = 0.29$

for MILK2006. The performance of PLSR is overall lower than the other models, and it is particularly poor at predicting DM yield, with a  $\tau_w$  value of only 0.60. RF exhibits consistent performance across each trait and excelled in predicting NDF, with  $\tau_w$  value of 0.42.

**Multi-Task Models:** Turning to deep learning models, 1D-CNN and Att-1D-CNN provide a comparable performance across all qualities. 1D-CNN leads with the highest  $\tau_w$  value of 0.71 for CP, followed by DNN with the score of 0.69. Bi-LSTM and att-bi-LSTM have a moderate performance across all qualities. From the results, while attention mechanisms do not consistently enhance all models, they do augment the predictive capacity of certain traits such as NDF and starch. MUSTA outperforms all the other models in most of the qualities. Particularly, it stands out for DM yield, ADF, starch, and MILK2006 with  $\tau_w$  of 0.79, 0.51, 0.42, and 0.74, making it the best choice for these qualities among the evaluated models.

In summary, the MUSTA models show the highest  $\tau_w$  for DM yield, ADF, starch, and MILK2006, while RF stands out for NDF and 1D-CNN stands out for CP.

#### 4.1.2. Model Classification Performance

The results in Appendix D Table A2 demonstrate the comparative assessment of six multi-task deep learning models applied to classification tasks. Specifically, the results show:

**DM yield:** In assessing the various models for their classification capabilities, MUSTA emerges as the most effective model for DM yield predictions. It delivers the highest  $F1_{top\_c}$  score at 0.70, an  $Accuracy_{avg}$  of 0.76, and a  $Precision_{top\_c}$  of 0.69. In a close competition, the att-1D-CNN model demonstrates performance with an  $F1_{top\_c}$  score of 0.69, an  $Accuracy_{avg}$  of 0.74, and a  $Precision_{top\_c}$  of 0.63. Moreover, att-1D-CNN outperforms other models in  $Recall_{top\_c}$  by obtaining a score of 0.77.

**NDF:** When estimating NDF, att-1D-CNN stands out by achieving the highest scores in  $F1_{top\_c}$ ,  $Accuracy_{avg}$ , and  $Recall_{top\_c}$ , scoring 0.50, 0.64, and 0.56, respectively. DNN excels in  $Precision_{top\_c}$  by scoring 0.49, the highest among the evaluated models.

**ADF:** In the context of ADF predictions, both MUSTA and 1D-CNN models display a good performance with  $F1_{top\_c}$  score of 0.51. The att-1D-CNN model shows the highest  $Accuracy_{avg}$  at 0.65, while MUSTA outperforms others in  $Recall_{top\_c}$ , achieving a score of 0.55.

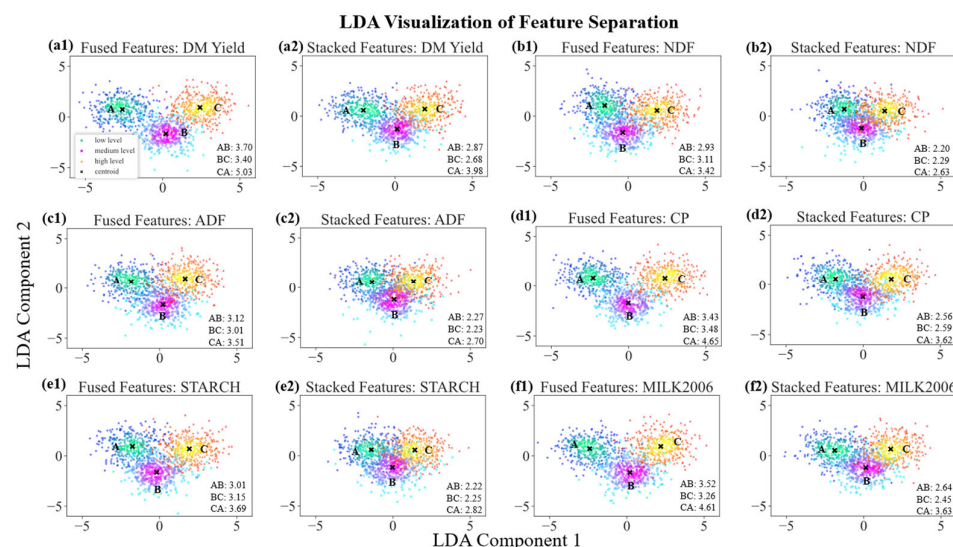
**CP:** For the CP estimations, 1D-CNN proves superior by leading in  $F1_{top\_c}$ ,  $Accuracy_{avg}$ , and  $Precision_{top\_c}$  with respective scores of 0.66, 0.72, and 0.66. MUSTA and DNN outperform others in  $Recall_{top\_c}$ , achieving a score of 0.67.

**Starch:** During starch predictions, Att-1D-CNN dominates by achieving the highest scores in  $F1_{top\_c}$ ,  $Accuracy_{avg}$ , and  $Recall_{top\_c}$  with respective scores of 0.54, 0.65, and 0.59. DNN outperforms others in  $Precision_{top}$  with a score of 0.52.

**MILK2006:** For the MILK2006 estimation, MUSTA stands out with the highest  $F1_{top\_c}$ ,  $Accuracy_{avg}$ , and  $Recall_{top\_c}$  of 0.65, 0.72, and 0.63, respectively. The highest score of  $Recall_{top\_c}$  was achieved by att-1D-CNN with a score of 0.74.

#### 4.2. Class Separability Analysis of Fused Features vs. Stacked Features

Figure 6 presents LDA visualization that compares the degree of class separation between fused and stacked features across all samples. From the projected scatter points in the LDA components, a noticeably greater degree of overlap is observed between the three classes for the stacked features, which means a worse separation between different classes, when compared with deep fused features.



**Figure 6.** LDA visualization illustrating the class separation on fused features vs. stacked features on estimating the respective traits. A, B, and C denote the centroids for the low, medium, and high quality value levels of the respective traits. (a1) Fused features: DM yield; (a2) Stacked features: DM yield; (b1) Fused features: NDF; (b2) Stacked features: NDF; (c1) Fused features: ADF; (c2) Stacked features: ADF; (d1) Fused features: CP; (d2) Stacked features: CP; (e1) Fused features: STARCH; (e2) Stacked features: STARCH; (f1) Fused features: MILK2006; (f2) Stacked features: MILK2006.

The overlap becomes more evident considering the relatively shorter centroid distances within the transformed LDA feature space for stacked features. These centroids give an average representation of each class in the LDA feature space. A greater centroid distance indicates better class separation. For instance, in the LDA feature space for MILK2006, deep fused features showed larger centroid distances among each category (AB: 3.52, BC: 3.26, CA: 4.61; A, B, and C denote the centroids for the low, middle, and high quality value levels of the respective trait) compared to the corresponding stacked features (AB: 2.64, BC: 2.45, CA: 3.63). We also conducted the Wilcoxon Signed-Rank test on all the category centroid distances yielded by the stacked and fused features, respectively, to validate the class separation observed in the LDA visualization. The results showed a  $p$ -value of  $7.63 \times 10^{-6}$ , which is significantly smaller than 0.05, indicating a statistically significant improvement in class separability with the fused features.

#### 4.3. Feature Importance Analysis

In this section, we aim to illustrate the significance of data modalities and data acquisition dates in relation to their respective influence on the estimation of various quality traits. By doing so, we seek to identify the most effective modality and optimal time of data collection for each distinct trait.

##### 4.3.1. Feature Importance on UAV Sensor Modalities

Figure 7 provides a comprehensive assessment of feature importance for various quality traits.

**DM yield:** The canopy hyperspectral features contribute the most (38%) to yield prediction. This is followed by LiDAR structural features (17%) and LiDAR intensity features (11%). Canopy RGB morphological features and canopy RGB textural features exhibit less importance, at 6% and 4%, respectively.

**NDF:** In predicting NDF, LiDAR structural features are the most important (38%). The canopy RGB textural features and canopy RGB morphological features follow with feature

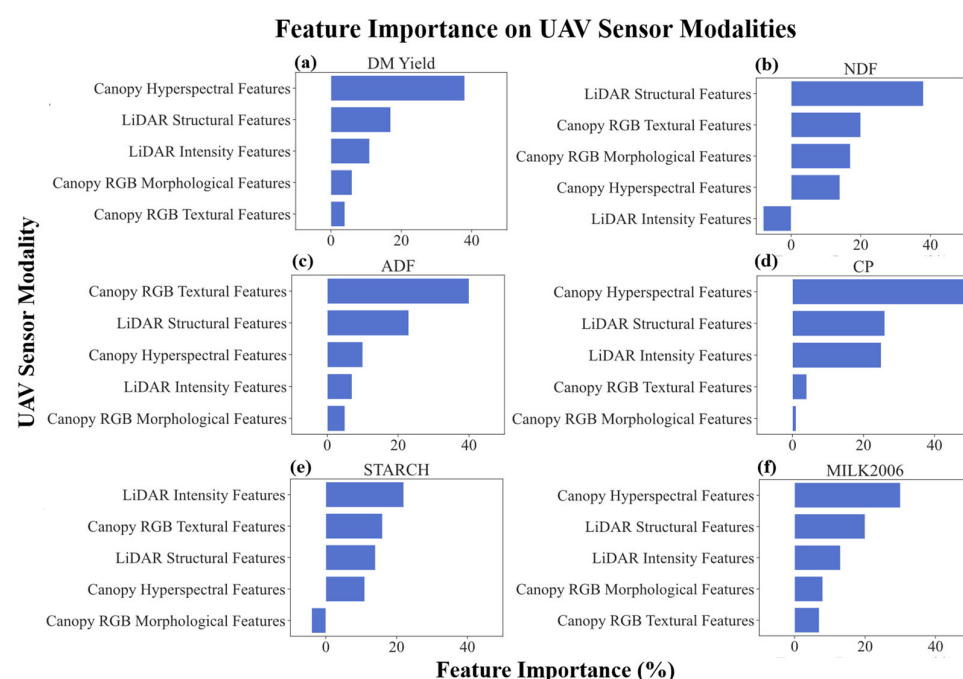
importance of 20% and 17%, respectively. The LiDAR intensity features hold negative importance (−8%) in predicting NDF.

**ADF:** For ADF, the canopy RGB textural features have the highest importance (40%). This is followed by the LiDAR structural features at 23%. The other features' importance is all below 10%.

**CP:** The prediction of CP is heavily influenced by canopy hyperspectral features (49%). The LiDAR features, both structural and intensity, also contribute significantly, at 26% and 25%, respectively.

**Starch:** In estimating starch, the LiDAR intensity features come out on top (22%), followed by canopy RGB textural features (16%) and LiDAR structural features (14%). Canopy RGB morphological features hold negative importance (−4%) in this case.

**MILK2006:** For MILK2006, canopy hyperspectral features are the most impactful (30%), followed by LiDAR structural features (20%). The other features exhibit similar levels of importance, ranging from 7% to 13%.



**Figure 7.** Feature importance of different UAV sensor modalities in estimating six quality traits. (a) DM yield; (b) NDF; (c) ADF; (d) CP; (e) starch; (f) MILK2006. Bars indicate the relative importance (%) of canopy hyperspectral, LiDAR structural, LiDAR intensity, canopy RGB textural, and canopy RGB morphological features for each trait.

#### 4.3.2. Feature Importance on UAV Survey Timing

Figure 8 shows the feature importance of different data acquisition dates, represented by days after sowing (DAS), in estimating six quality traits.

**DM yield:** The results indicate that features obtained 108 DAS are most impactful (32%) in yield prediction. The next most influential features are those acquired 53 DAS (15%), while the rest show relatively less importance, all scoring below 10%.

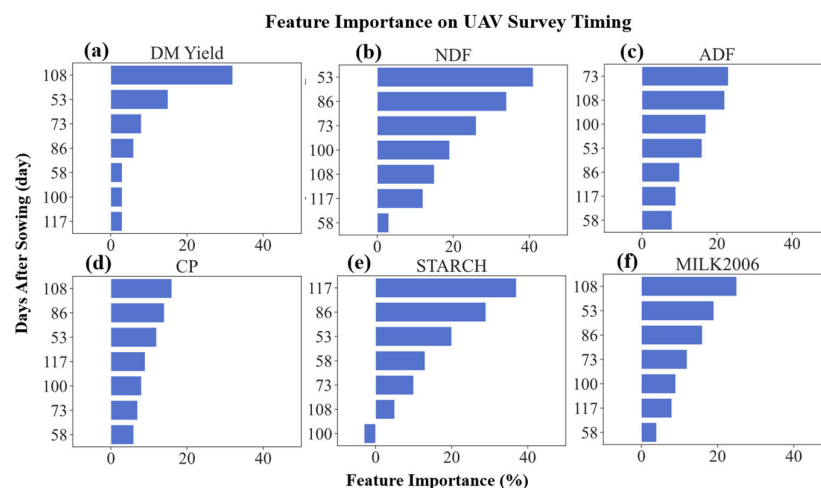
**NDF:** For NDF prediction, the features obtained 53 DAS (41%) and 86 DAS (34%) are the most important. The other days' features have less impact, with the feature importance scores ranging from 3% to 26%.

**ADF:** In terms of ADF prediction, the data acquired 73 DAS (23%) and 108 DAS (22%) play the most critical roles. Features from other days are less influential, with their importance scores varying between 8% and 16%.

**CP:** For predicting CP, the features derived 108 DAS (16%) hold the highest importance, followed by those captured 86 DAS (14%). The rest of the days offer less influence, all featuring scores around 10%.

**Starch:** When estimating starch content, the data obtained 117 DAS (37%) and 86 DAS (29%) are the most significant. Notably, the data captured 100 DAS exhibits a negative feature importance score (−3%).

**MILK2006:** For MILK2006, the feature importance peak is observed for the data obtained 108 DAS (25%), followed by data obtained 53 DAS (19%) and 86 DAS (16%). The remaining dates contribute less to the prediction of this trait.



**Figure 8.** Feature importance of different UAV survey timing in estimating six quality traits. (a) DM yield; (b) NDF; (c) ADF; (d) CP; (e) starch; (f) MILK2006. Bars indicate the relative importance (%) of features derived from each survey date for the corresponding trait.

#### 4.4. Retrieval-Augmented Method Performance

Table 5 illustrates the regression performance of the retrieval-based approach used to predict maize silage yield and quality measurements. Table 6 depicts the efficacy metrics of the retrieval-based method when applied to the classification of yield and quality levels in maize silage. The percentages given in parentheses indicate the relative performance ratio (RPR) of the retrieval-based method in comparison to the original model. Specifically, this ratio is given by Equation (14).

$$\text{Relative Performance Ratio} = \frac{\text{Retrieval} - \text{Based Method Performance}}{\text{Original Performance}} \times 100\% \quad (14)$$

**Table 5.** Performance of the retrieval-augmented method in estimating the maize quality values. The percentages given in parentheses indicate the relative performance ratio.

| Quality Values | Metrics     |            |             |             |
|----------------|-------------|------------|-------------|-------------|
|                | $\tau_w$    | $r$        | MAE         | RMSE        |
| DM Yield       | 0.67 (85%)  | 0.72 (88%) | 0.92 (121%) | 1.16 (122%) |
| NDF            | 0.41 (105%) | 0.33 (94%) | 1.90 (101%) | 2.39 (100%) |
| ADF            | 0.49 (96%)  | 0.38 (93%) | 1.42 (102%) | 1.78 (101%) |
| CP             | 0.55 (81%)  | 0.58 (85%) | 0.39 (111%) | 0.49 (111%) |
| Starch         | 0.43 (102%) | 0.38 (95%) | 2.29 (100%) | 2.88 (100%) |
| MILK2006       | 0.67 (91%)  | 0.69 (91%) | 1.59 (110%) | 2.00 (111%) |

It is noteworthy that the retrieval-augmented method performs well by retrieving the costly hyperspectral features from the training database and then making predictions. For many traits, this method displays a performance in regression and classification that is

comparable to, or in certain less predictable traits, even superior to, the full-featured model. This is particularly evident in the regression for traits such as NDF (RPR on  $\tau_w$ : 105%) and starch (RPR on  $\tau_w$ : 102%).

**Table 6.** Performance of the retrieval-augmented method in classifying the maize silage quality levels. The percentages given in parentheses indicate the relative performance ratio.

| Quality Values | Metrics       |                  |                      |                   |
|----------------|---------------|------------------|----------------------|-------------------|
|                | $F1_{top\_c}$ | $Accuracy_{avg}$ | $Precision_{top\_c}$ | $Recall_{top\_c}$ |
| DM Yield       | 0.64 (91%)    | 0.72 (95%)       | 0.64 (93%)           | 0.64 (89%)        |
| NDF            | 0.45 (95%)    | 0.61 (100%)      | 0.43 (100%)          | 0.46 (90%)        |
| ADF            | 0.49 (96%)    | 0.64 (100%)      | 0.47 (98%)           | 0.50 (91%)        |
| CP             | 0.57 (89%)    | 0.67 (96%)       | 0.56 (92%)           | 0.58 (87%)        |
| Starch         | 0.47 (90%)    | 0.62 (97%)       | 0.47 (94%)           | 0.47 (87%)        |
| MILK2006       | 0.61 (94%)    | 0.70 (97%)       | 0.60 (95%)           | 0.61 (91%)        |

## 5. Discussion

### 5.1. Advantages of MUSTA

Experimental trials and validation processes above have shown that MUSTA provides key improvements in effectiveness and efficiency in the estimation of maize silage nutritive values. Among the assessed machine learning and deep learning models, MUSTA consistently surpasses the others in performance for simultaneously estimating most nutritive aspects. In MUSTA, att-1D-CNN and att-bi-LSTM were able to uncover patterns behind the time-series data and then extract comprehensive crop growth status information. In regression tasks, MUSTA exhibited great potential in forecasting DM yield, ADF, starch, and MILK2006, as indicated by higher  $\tau_w$  values, suggesting a robust positive ranking correlation. Moreover, in classification tasks, the superiority of the MUSTA model became evident when examining the  $F1_{top\_c}$ ,  $Accuracy_{avg}$ ,  $Precision_{top\_c}$  and  $Recall_{top\_c}$  scores for different nutritive value levels. Notably, MUSTA showed the top  $F1_{top\_c}$  score,  $Accuracy_{avg}$ , and  $Precision_{top\_c}$  scores for DM yield and ultimate breeding criteria of MILK2006, confirming its effectiveness and efficiency. In addition, the LDA plots show that deep fused features, a crucial aspect of the MUSTA model, have greater cluster centroid distances in the transformed LDA feature space compared to stacked features. This suggests that MUSTA achieved superior class separation, thereby endorsing the model's predictive ability and efficiency.

### 5.2. Feature Importance Analysis

#### 5.2.1. UAV Sensor Modality Contribution

For DM yield prediction, canopy hyperspectral features contribute the most (38%). This could be due to the hyperspectral sensors capturing extensive information about the light that interacts with the crops, providing richer data that aids in yield prediction. Similarly, for CP and MILK2006, the dominance of hyperspectral features (49% and 30%, respectively) is observed. It demonstrates the essential role of hyperspectral imaging in capturing nutrient-related information. LiDAR structural features play a crucial role in predicting NDF (38%) and substantially contribute to DM yield (17%), CP (26%), and MILK2006 (20%). LiDAR technology enables precise three-dimensional measurements of plant structure, providing valuable information for predicting traits related to plant structure and nutritional content. Notably, LiDAR intensity features and canopy RGB morphological features exhibit a negative impact on predicting NDF and starch, respectively. This could suggest that these types of features might have great variability in the hybrids and then the model might be prone to overfitting these features when estimating these specific quality traits.

### 5.2.2. UAV Survey Timing Contribution

We also showcased the impact of data acquisition dates, represented as days after sowing (DAS). Notably, features acquired 108 DAS are the most impactful in predicting DM yield (32%), MILK2006 (25%), ADF (22%), and CP (16%). This suggests that the period around 108 DAS may be a critical growth stage where key characteristics determining these traits emerge. Silage maize undergoes various growth stages including germination, vegetative growth, flowering, grain fill, and maturity. The timing of approximately 108 DAS typically aligns with a phase of relative maturity, where grain development and stalk maturation occur [62]. At this stage, the plant might reach its fiber and DM content peak. This is due to the declining leaf-to-stem ratio (increased stem, fewer leaves) as the forage matures, which in turn reduces digestibility as a larger portion of the total fiber concentration is associated with stem tissue [63,64]. Conversely, data captured on certain days exhibits negative feature importance for some traits (e.g., starch at 100 DAS), indicating these times might not be the best for capturing relevant information for these specific traits.

### 5.3. Advantages of Retrieval-Based Method

The retrieval-based method demonstrated impressive results in predicting maize silage yield and quality measurements, offering several notable advantages. The method stands out because it retrieves costly hyperspectral features from the training dataset for predictions. This procedure might enable the model to utilize valuable and rich signals from the historical data without relying much on collecting new hyperspectral data for the next season or a new location, thereby reducing cost and time. This is particularly beneficial in tight-budget projects, where collecting and processing hyperspectral data can be expensive. The effectiveness of the retrieval-based approach becomes apparent when comparing its performance with the original model that incorporates all features. This method displayed a performance level in regression and classification that is on par with, or in some less predictable traits, even superior to the full-featured model. This efficiency is particularly visible in the regression for traits such as NDF (RPR on  $\tau_w$ : 105%) and starch (RPR on  $\tau_w$ : 102%). The exceptional performance of the retrieval-based method for traits that were typically harder to predict (e.g., NDF, ADF and starch) suggested it might have the ability to discern infrequent feature distributions caused by the unseen data in the test dataset that the original model struggled with. This could be because the retrieval-augmented method is particularly adept at bypassing the influence of these infrequent data distributions by retrieving and fusing the features of the nearest neighbors of the test samples from the training database [65], hence enhancing the model's proficiency in predicting more intricate traits.

### 5.4. Limitations and Future Work

One key limitation of this study is the model's less satisfactory performance in predicting NDF, ADF, and starch. At the model level, we also acknowledge that MUSTA underperforms the 1D-CNN and Att-1D-CNN baselines in the classification branch for NDF, CP, and starch. NDF and starch are more closely tied to cell-wall composition, lignification, and kernel development, which are only indirectly reflected in the 400–1000 nm spectral regions. Moreover, these traits are strongly modulated by genotype  $\times$  environment ( $G \times E$ ) and by maturity stage, so plots with similar canopy spectra can still diverge in fiber and starch concentrations. Researchers suggest that longer SWIR bands' feature would likely improve model predictability [16]. Another limitation is the system-level complexity, as the workflow require access to a multi-sensor UAV platform, accurate GNSS/IMU processing, and GPU-based deep learning models. Admittedly, the retrieval-augmented

method is a step toward reducing this dependence but still does not eliminate this system-level complexity. Finally, although sensor-level feature contributions were reported, we acknowledge the internal decision process of the deep learning model is still not transparent. Future work should incorporate more interpretable analyses and test models across broader environments to improve generalizability.

## 6. Conclusions

This study presents MUSTA, a model that merges multi-sensory features, built upon the principles of multi-task learning and attention mechanisms. The model is designed to concurrently estimate various nutritional attributes for silage maize hybrids. In this study, we extracted time-series features from multiple sensors during the seven-day UAV survey to characterize the crop growth dynamics throughout the growing season. Additionally, we integrated the attention mechanism to help alleviate the impact of feature imbalance by enabling the model to learn to focus on the most relevant ones. Moreover, we proposed a retrieval-based method enabling phenotype estimation, even in the absence of costly hyperspectral features, presenting a budget-friendly alternative for breeding or farming missions. Through experimentation and validation, we demonstrated that our proposed approach outperforms baselines of traditional single-task machine learning models and multi-task deep learning models in terms of efficiency and overall effectiveness. The findings of this study hold potential for shaping agricultural management practices and guiding future crop breeding programs, fostering an environment for accelerating genomic selection and developing next-generation crops.

**Author Contributions:** J.F.: Conceptualization, Software, Investigation, Data Curation, Methodology, Formal Analysis, Visualization, Validation, Writing—Original Draft. J.Z.: Conceptualization, Investigation, Validation, Writing—Review and Editing. N.d.L.: Resources, Data Curation, Validation, Writing—Review and Editing. Z.Z.: Supervision, Resources, Conceptualization, Validation, Writing—Review and Editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture, Agriculture and Food Research Initiative Foundational Program (Award No. 2022-67021-36469), and Wisconsin Dairy Innovation Hub.

**Data Availability Statement:** Aerial data this study are not publicly available due to privacy and administrative restrictions. Additional agricultural field data and maize breeding genomic data from the Genomes to Fields (G2F) Initiative and past University of Wisconsin–Madison breeding experiments are available through the G2F Resources website: <https://www.genomes2fields.org/resources> (accessed on 1 September 2025).

**Acknowledgments:** We acknowledge the USDA Germplasm repository and the Germplasm Enhancement of Maize program for providing useful germplasm.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A

- RGB Texture Features (20 features): used skimage Python library to compute the GLCM at a distance of 1 pixel in four directions (0°, 45°, 90°, 135°). From these matrices, five properties are calculated: Contrast, Correlation, Energy, Homogeneity, and Dissimilarity, resulting in 20 texture features per plot.
- RGB Morphological Features (25 features): used the uib\_vfeatures Python library to automatically calculate 25 shape features from the cleaned canopy mask [50]. The features include Solidity, CH Perimeter, CH Area, BB Area, Rectangularity, Min r, Max

r, Feret, Breadth, Circularity, Roundness, Feret Angle, Eccentricity, Center, Sphericity, Aspect Ratio, Area equivalent, Perimeter equivalent, Equivalent ellipse area, Compactness, Area, Convexity, Shape, Perimeter, Bounding\_box\_area, and Shape Factor—a total of 25 features that quantify the shape of the canopy.

## Appendix B

- LiDAR Structural Features (36 features)
  - a. The 10th through 99th percentiles of relative plant height (10 features).
  - b. Statistical moments describing the distribution of relative heights (5 features): standard\_deviation, quadratic\_mean, skewness, kurtosis, variation.
  - c. Canopy Cover (7 features): Estimated as the ratio of canopy points to the total number of points above seven different height thresholds (0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.75 multiplied by the 99th percentile plot height) [66].
  - d. Canopy Volume (1 feature): The plot is divided into a grid of  $8 \times 8$  cm cells. The volume is the sum of all cell volumes, where each cell's volume is its  $8 \times 8$  cm area multiplied by the 95th percentile relative height of points within it [66].
  - e. Projected Leaf Area (PLA) (7 features): Calculated similarly to canopy cover but multiplied by the cell resolution and total grid area to estimate an aggregated area [53].
  - f. Plant Area Index (PAI) (1 feature): Calculated from voxelized LiDAR data using the voxel-based canopy profiling method proposed by Hosoi and Omasa, which provides an estimate of the total one-sided plant area per unit ground area [53].
  - g. Plant Area Density (PAD) (5 features): Describes the vertical distribution of plant material within the canopy. Our study calculated PAD for vertical layers (from 0 to 4 m with a 10 cm resolution), providing a detailed profile of canopy structure [53].
- LiDAR Intensity Features (16 features)
  - a. The 10th through 99th percentiles of point intensity values (10 features).
  - b. Statistical moments describing the distribution of point intensity (5 features): standard\_deviation, quadratic\_mean, skewness, kurtosis, variation.
  - c. Point Cloud Statistics (1 features): total non-ground points.

## Appendix C

**Table A1.** Performance of five single-task traditional machine learning models and six deep-learning-based multi-task models in estimating the maize silage yield and nutritive values.

| Quality Values            |                       | Models        | Metrics     |      |      |      |
|---------------------------|-----------------------|---------------|-------------|------|------|------|
|                           |                       |               | $\tau_w$    | $r$  | MAE  | RMSE |
| DM Yield<br>(US ton/acre) | Single-Task<br>Models | Ridge         | 0.75        | 0.79 | 0.84 | 1.07 |
|                           |                       | LASSO         | 0.73        | 0.78 | 0.82 | 1.06 |
|                           |                       | SVR           | 0.65        | 0.72 | 0.93 | 1.20 |
|                           |                       | PLSR          | 0.60        | 0.68 | 0.95 | 1.22 |
|                           |                       | Random Forest | 0.70        | 0.73 | 0.90 | 1.15 |
|                           | Multi-Task<br>Models  | DNN           | 0.74        | 0.78 | 0.82 | 1.05 |
|                           |                       | 1D-CNN        | 0.75        | 0.80 | 0.79 | 1.01 |
|                           |                       | Att-1D-CNN    | 0.75        | 0.79 | 0.82 | 1.04 |
|                           |                       | Bi-LSTM       | 0.70        | 0.74 | 0.92 | 1.16 |
|                           |                       | Att-Bi-LSTM   | 0.70        | 0.77 | 0.87 | 1.09 |
|                           |                       | MUSTA         | <b>0.79</b> | 0.82 | 0.76 | 0.95 |

Table A1. Cont.

| Quality Values         |                    | Models        | Metrics     |      |      |      |
|------------------------|--------------------|---------------|-------------|------|------|------|
|                        |                    |               | $\tau_w$    | $r$  | MAE  | RMSE |
| NDF (%)                | Single-Task Models | Ridge         | 0.29        | 0.33 | 2.04 | 2.65 |
|                        |                    | LASSO         | 0.31        | 0.29 | 1.91 | 2.39 |
|                        |                    | SVR           | 0.30        | 0.23 | 1.95 | 2.44 |
|                        |                    | PLSR          | 0.21        | 0.21 | 1.95 | 2.44 |
|                        |                    | Random Forest | <b>0.42</b> | 0.36 | 1.86 | 2.34 |
|                        | Multi-Task Models  | DNN           | 0.37        | 0.32 | 2.15 | 2.71 |
|                        |                    | 1D-CNN        | 0.29        | 0.30 | 2.03 | 2.55 |
|                        |                    | Att-1D-CNN    | 0.35        | 0.34 | 1.90 | 2.41 |
|                        |                    | Bi-LSTM       | 0.36        | 0.32 | 2.03 | 2.53 |
|                        |                    | Att-Bi-LSTM   | 0.38        | 0.32 | 2.00 | 2.50 |
|                        |                    | MUSTA         | 0.39        | 0.35 | 1.89 | 2.38 |
| ADF (%)                | Single-Task Models | Ridge         | 0.39        | 0.38 | 1.50 | 1.95 |
|                        |                    | LASSO         | 0.47        | 0.38 | 1.40 | 1.76 |
|                        |                    | SVR           | 0.36        | 0.33 | 1.43 | 1.80 |
|                        |                    | PLSR          | 0.36        | 0.30 | 1.44 | 1.81 |
|                        |                    | Random Forest | 0.47        | 0.41 | 1.39 | 1.74 |
|                        | Multi-Task Models  | DNN           | 0.45        | 0.35 | 1.60 | 2.02 |
|                        |                    | 1D-CNN        | 0.42        | 0.36 | 1.47 | 1.86 |
|                        |                    | Att-1D-CNN    | 0.40        | 0.40 | 1.42 | 1.79 |
|                        |                    | Bi-LSTM       | 0.37        | 0.36 | 1.51 | 1.89 |
|                        |                    | Att-Bi-LSTM   | 0.41        | 0.34 | 1.50 | 1.88 |
|                        |                    | MUSTA         | <b>0.51</b> | 0.41 | 1.39 | 1.76 |
| CP (%)                 | Single-Task Models | Ridge         | 0.60        | 0.67 | 0.36 | 0.47 |
|                        |                    | LASSO         | 0.62        | 0.65 | 0.38 | 0.48 |
|                        |                    | SVR           | 0.64        | 0.67 | 0.36 | 0.46 |
|                        |                    | PLSR          | 0.53        | 0.57 | 0.40 | 0.49 |
|                        |                    | Random Forest | 0.64        | 0.65 | 0.37 | 0.47 |
|                        | Multi-Task Models  | DNN           | 0.69        | 0.70 | 0.35 | 0.44 |
|                        |                    | 1D-CNN        | <b>0.71</b> | 0.71 | 0.34 | 0.42 |
|                        |                    | Att-1D-CNN    | 0.68        | 0.69 | 0.35 | 0.44 |
|                        |                    | Bi-LSTM       | 0.62        | 0.67 | 0.36 | 0.45 |
|                        |                    | Att-Bi-LSTM   | 0.60        | 0.66 | 0.37 | 0.46 |
|                        |                    | MUSTA         | 0.68        | 0.68 | 0.35 | 0.44 |
| STARCH (%)             | Single-Task Models | Ridge         | 0.34        | 0.40 | 2.39 | 3.09 |
|                        |                    | LASSO         | 0.38        | 0.35 | 2.31 | 2.89 |
|                        |                    | SVR           | 0.34        | 0.32 | 2.35 | 2.93 |
|                        |                    | PLSR          | 0.27        | 0.29 | 2.37 | 2.95 |
|                        |                    | Random Forest | 0.37        | 0.40 | 2.25 | 2.83 |
|                        | Multi-Task Models  | DNN           | 0.40        | 0.36 | 2.55 | 3.23 |
|                        |                    | 1D-CNN        | 0.38        | 0.37 | 2.37 | 3.00 |
|                        |                    | Att-1D-CNN    | 0.42        | 0.42 | 2.26 | 2.86 |
|                        |                    | Bi-LSTM       | 0.40        | 0.39 | 2.37 | 3.01 |
|                        |                    | Att-Bi-LSTM   | <b>0.42</b> | 0.37 | 2.35 | 2.96 |
|                        |                    | MUSTA         | <b>0.42</b> | 0.40 | 2.28 | 2.89 |
| MILK2006 (US ton/acre) | Single-Task Models | Ridge         | 0.63        | 0.71 | 1.61 | 2.04 |
|                        |                    | LASSO         | 0.71        | 0.74 | 1.48 | 1.85 |
|                        |                    | SVR           | 0.29        | 0.26 | 2.12 | 2.75 |
|                        |                    | PLSR          | 0.52        | 0.63 | 1.67 | 2.15 |
|                        |                    | Random Forest | 0.63        | 0.67 | 1.61 | 2.06 |
|                        | Multi-Task Models  | DNN           | 0.65        | 0.70 | 1.60 | 2.02 |
|                        |                    | 1D-CNN        | 0.68        | 0.72 | 1.52 | 1.91 |
|                        |                    | Att-1D-CNN    | 0.65        | 0.72 | 1.54 | 1.94 |
|                        |                    | Bi-LSTM       | 0.60        | 0.67 | 1.67 | 2.10 |
|                        |                    | Att-Bi-LSTM   | 0.61        | 0.69 | 1.58 | 2.01 |
|                        |                    | MUSTA         | <b>0.74</b> | 0.76 | 1.45 | 1.80 |

Values in bold underlined indicate the best performance for each metric.

## Appendix D

**Table A2.** Performance of six deep-learning-based multi-task models in classifying the maize silage yield and nutritive levels.

| Quality Values | Models            | Metrics       |                  |                      |                   |             |
|----------------|-------------------|---------------|------------------|----------------------|-------------------|-------------|
|                |                   | $F1_{top\_c}$ | $Accuracy_{avg}$ | $Precision_{top\_c}$ | $Recall_{top\_c}$ |             |
| DM Yield       | Multi-Task Models | DNN           | 0.68             | 0.74                 | 0.66              | 0.70        |
|                |                   | 1D-CNN        | 0.64             | 0.73                 | 0.64              | 0.63        |
|                |                   | Att-1D-CNN    | 0.69             | 0.74                 | 0.63              | <u>0.77</u> |
|                |                   | Bi-LSTM       | 0.65             | 0.72                 | 0.58              | 0.75        |
|                |                   | Att-Bi-LSTM   | 0.68             | 0.73                 | 0.62              | 0.75        |
|                |                   | MUSTA         | <u>0.70</u>      | <u>0.76</u>          | <u>0.69</u>       | 0.72        |
| NDF            | Multi-Task Models | DNN           | 0.39             | 0.62                 | <u>0.49</u>       | 0.33        |
|                |                   | 1D-CNN        | 0.49             | 0.62                 | 0.46              | 0.52        |
|                |                   | Att-1D-CNN    | <u>0.50</u>      | <u>0.64</u>          | 0.46              | <u>0.56</u> |
|                |                   | Bi-LSTM       | 0.44             | 0.62                 | 0.46              | 0.42        |
|                |                   | Att-Bi-LSTM   | 0.44             | 0.61                 | 0.45              | 0.43        |
|                |                   | MUSTA         | 0.47             | 0.61                 | 0.43              | 0.51        |
| ADF            | Multi-Task Models | DNN           | 0.41             | 0.62                 | <u>0.51</u>       | 0.34        |
|                |                   | 1D-CNN        | <u>0.51</u>      | 0.63                 | 0.49              | 0.54        |
|                |                   | Att-1D-CNN    | 0.50             | <u>0.65</u>          | 0.46              | 0.54        |
|                |                   | Bi-LSTM       | 0.47             | 0.63                 | 0.50              | 0.45        |
|                |                   | Att-Bi-LSTM   | 0.45             | 0.61                 | 0.46              | 0.43        |
|                |                   | MUSTA         | <u>0.51</u>      | 0.64                 | 0.48              | <u>0.55</u> |
| CP             | Multi-Task Models | DNN           | 0.64             | 0.70                 | 0.61              | <u>0.67</u> |
|                |                   | 1D-CNN        | <u>0.66</u>      | <u>0.72</u>          | <u>0.66</u>       | 0.65        |
|                |                   | Att-1D-CNN    | 0.62             | 0.70                 | 0.61              | 0.62        |
|                |                   | Bi-LSTM       | 0.63             | 0.70                 | 0.61              | 0.66        |
|                |                   | Att-Bi-LSTM   | 0.61             | 0.70                 | 0.64              | 0.58        |
|                |                   | MUSTA         | 0.64             | 0.70                 | 0.61              | <u>0.67</u> |
| STARCH         | Multi-Task Models | DNN           | 0.44             | 0.63                 | <u>0.52</u>       | 0.38        |
|                |                   | 1D-CNN        | 0.53             | 0.64                 | 0.49              | 0.57        |
|                |                   | Att-1D-CNN    | <u>0.54</u>      | <u>0.65</u>          | 0.49              | <u>0.59</u> |
|                |                   | Bi-LSTM       | 0.53             | 0.64                 | 0.51              | 0.54        |
|                |                   | Att-Bi-LSTM   | 0.47             | 0.62                 | 0.48              | 0.47        |
|                |                   | MUSTA         | 0.52             | 0.64                 | 0.50              | 0.54        |
| MILK2006       | Multi-Task Models | DNN           | 0.59             | 0.70                 | 0.62              | 0.55        |
|                |                   | 1D-CNN        | 0.59             | 0.69                 | 0.59              | 0.58        |
|                |                   | Att-1D-CNN    | 0.64             | 0.70                 | 0.56              | <u>0.74</u> |
|                |                   | Bi-LSTM       | 0.61             | 0.69                 | 0.54              | 0.69        |
|                |                   | Att-Bi-LSTM   | 0.61             | 0.70                 | 0.56              | 0.67        |
|                |                   | MUSTA         | <u>0.65</u>      | <u>0.72</u>          | <u>0.63</u>       | 0.67        |

Values in bold underlined indicate the best performance for each metric.

## References

1. Zhao, M.; Feng, Y.; Shi, Y.; Shen, H.; Hu, H.; Luo, Y.; Xu, L.; Kang, J.; Xing, A.; Wang, S.; et al. Yield and Quality Properties of Silage Maize and Their Influencing Factors in China. *Sci. China Life Sci.* **2022**, *65*, 1655–1666. [[CrossRef](#)] [[PubMed](#)]
2. Martin, N.P.; Russelle, M.P.; Powell, J.M.; Sniffen, C.J.; Smith, S.I.; Tricarico, J.M.; Grant, R.J. Invited Review: Sustainable Forage and Grain Crop Production for the US Dairy Industry. *J. Dairy Sci.* **2017**, *100*, 9479–9494. [[CrossRef](#)] [[PubMed](#)]
3. Stevenson, J.R.; Villoria, N.; Byerlee, D.; Kelley, T.; Maredia, M. Green Revolution Research Saved an Estimated 18 to 27 Million Hectares from Being Brought into Agricultural Production. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8363–8368. [[CrossRef](#)] [[PubMed](#)]
4. Bornowski, N.; Michel, K.J.; Hamilton, J.P.; Ou, S.; Seetharam, A.S.; Jenkins, J.; Grimwood, J.; Plott, C.; Shu, S.; Talag, J.; et al. Genomic Variation within the Maize Stiff-Stalk Heterotic Germplasm Pool. *Plant Genome* **2021**, *14*, e20114. [[CrossRef](#)]
5. Jiang, M.; Ma, Y.; Khan, N.; Khan, M.Z.; Akbar, A.; Khan, R.U.; Kamran, M.; Khan, N.A. Effect of Spring Maize Genotypes on Fermentation and Nutritional Value of Whole Plant Maize Silage in Northern Pakistan. *Fermentation* **2022**, *8*, 587. [[CrossRef](#)]
6. Perisic, M.; Perkins, A.; Lima, D.C.; de Leon, N.; Mitrovic, B.; Stanisavljevic, D. GEM Project-Derived Maize Lines Crossed with Temperate Elite Tester Lines Make for High-Quality, High-Yielding and Stable Silage Hybrids. *Agronomy* **2023**, *13*, 243. [[CrossRef](#)]
7. Johnson, L.M.; Harrison, J.H.; Davidson, D.; Robutti, J.L.; Swift, M.; Mahanna, W.C.; Shinnors, K. Corn Silage Management I: Effects of Hybrid, Maturity, and Mechanical Processing on Chemical and Physical Characteristics. *J. Dairy Sci.* **2002**, *85*, 833–853. [[CrossRef](#)]

8. Lorenz, A.J.; Beissinger, T.M.; Silva, R.R.; de Leon, N. Selection for Silage Yield and Composition Did Not Affect Genomic Diversity Within the Wisconsin Quality Synthetic Maize Population. *G3 Genes Genomes Genet.* **2015**, *5*, 541–549. [[CrossRef](#)]
9. Furbank, R.T.; Tester, M. Phenomics—Technologies to Relieve the Phenotyping Bottleneck. *Trends Plant Sci.* **2011**, *16*, 635–644. [[CrossRef](#)]
10. Kung, L.; Shaver, R.D.; Grant, R.J.; Schmidt, R.J. Silage Review: Interpretation of Chemical, Microbial, and Organoleptic Components of Silages. *J. Dairy Sci.* **2018**, *101*, 4020–4033. [[CrossRef](#)]
11. Buxton, D.R.; Muck, R.E.; Harrison, J.H. *Silage Science and Technology*; American Society of Agronomy, Inc.: Madison, WI, USA, 2015; ISBN 9780891182344.
12. Cherney, J.H.; Parsons, D.; Cherney, D.J.R. A Method for Forage Yield and Quality Assessment of Tall Fescue Cultivars in the Spring. *Crop Sci.* **2011**, *51*, 2878–2885. [[CrossRef](#)]
13. Norris, K.H.; Barnes, R.F.; Moore, J.E.; Shenk, J.S. Predicting Forage Quality by Infrared Reflectance Spectroscopy. *J. Anim. Sci.* **1976**, *43*, 889–897. [[CrossRef](#)]
14. Varela, J.I.; Miller, N.D.; Infante, V.; Kaeppler, S.M.; de Leon, N.; Spalding, E.P. A Novel High-Throughput Hyperspectral Scanner and Analytical Methods for Predicting Maize Kernel Composition and Physical Traits. *Food Chem.* **2022**, *391*, 133264. [[CrossRef](#)] [[PubMed](#)]
15. Starks, P.J.; Zhao, D.; Phillips, W.A.; Coleman, S.W. Development of Canopy Reflectance Algorithms for Real-Time Prediction of Bermudagrass Pasture Biomass and Nutritive Values. *Crop Sci.* **2006**, *46*, 927–934. [[CrossRef](#)]
16. Hossain, M.E.; Kabir, M.A.; Zheng, L.; Swain, D.L.; McGrath, S.; Medway, J. Near-Infrared Spectroscopy for Analysing Livestock Diet Quality: A Systematic Review. *Heliyon* **2024**, *10*, e40016. [[CrossRef](#)]
17. Hu, C.; Zhao, T.; Duan, Y.; Zhang, Y.; Wang, X.; Li, J.; Zhang, G. Visible-near Infrared Hyperspectral Imaging for Non-Destructive Estimation of Leaf Nitrogen Content under Water-Saving Irrigation in Protected Tomato Cultivation. *Front. Plant Sci.* **2025**, *16*, 1676457. [[CrossRef](#)]
18. Geipel, J.; Bakken, A.K.; Jørgensen, M.; Korsæth, A. Forage Yield and Quality Estimation by Means of UAV and Hyperspectral Imaging. *Precis. Agric.* **2021**, *22*, 1437–1463. [[CrossRef](#)]
19. Feng, L.; Zhang, Z.; Ma, Y.; Sun, Y.; Du, Q.; Williams, P.; Drewry, J.; Luck, B. Multitask Learning of Alfalfa Nutritive Value From UAV-Based Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 5506305. [[CrossRef](#)]
20. Feng, L.; Zhang, Z.; Ma, Y.; Du, Q.; Williams, P.; Drewry, J.; Luck, B. Alfalfa Yield Prediction Using UAV-Based Hyperspectral Imagery and Ensemble Learning. *Remote Sens.* **2020**, *12*, 2028. [[CrossRef](#)]
21. Hörtensteiner, S.; Matile, P. How Leaves Turn Yellow: Catabolism of Chlorophyll. In *Plant Cell Death Processes*; Academic Press: Cambridge, MA, USA, 2003; pp. 189–202. [[CrossRef](#)]
22. Maimaitijiang, M.; Ghulam, A.; Sidike, P.; Hartling, S.; Maimaitiyiming, M.; Peterson, K.; Shavers, E.; Fishman, J.; Peterson, J.; Kadam, S.; et al. Unmanned Aerial System (UAS)-Based Phenotyping of Soybean Using Multi-Sensor Data Fusion and Extreme Learning Machine. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 43–58. [[CrossRef](#)]
23. Feng, A.; Zhou, J.; Vories, E.D.; Sudduth, K.A.; Zhang, M. Yield Estimation in Cotton Using UAV-Based Multi-Sensor Imagery. *Biosyst. Eng.* **2020**, *193*, 101–114. [[CrossRef](#)]
24. Zhang, M.; Zhou, J.; Sudduth, K.A.; Kitchen, N.R. Estimation of Maize Yield and Effects of Variable-Rate Nitrogen Application Using UAV-Based RGB Imagery. *Biosyst. Eng.* **2020**, *189*, 24–35. [[CrossRef](#)]
25. Niu, Y.; Zhang, L.; Zhang, H.; Han, W.; Peng, X. Estimating Above-Ground Biomass of Maize Using Features Derived from UAV-Based RGB Imagery. *Remote Sens.* **2019**, *11*, 1261. [[CrossRef](#)]
26. Lu, J.; Cheng, D.; Geng, C.; Zhang, Z.; Xiang, Y.; Hu, T. Combining Plant Height, Canopy Coverage and Vegetation Index from UAV-Based RGB Images to Estimate Leaf Nitrogen Concentration of Summer Maize. *Biosyst. Eng.* **2021**, *202*, 42–54. [[CrossRef](#)]
27. Zhang, X.; Zhang, K.; Sun, Y.; Zhao, Y.; Zhuang, H.; Ban, W.; Chen, Y.; Fu, E.; Chen, S.; Liu, J.; et al. Combining Spectral and Texture Features of UAS-Based Multispectral Images for Maize Leaf Area Index Estimation. *Remote Sens.* **2022**, *14*, 331. [[CrossRef](#)]
28. Han, W.; Sun, Y.; Xu, T.; Chen, X.; Su, K.O. Detecting Maize Leaf Water Status by Using Digital RGB Images. *Int. J. Agric. Biol. Eng.* **2014**, *7*, 45–53. [[CrossRef](#)]
29. Lang, Q.; Zhiyong, Z.; Longsheng, C.; Hong, S.; Minzan, L.; Li, L.; Junyong, M. Detection of Chlorophyll Content in Maize Canopy from UAV Imagery. *IFAC-Pap.* **2019**, *52*, 330–335. [[CrossRef](#)]
30. Shrestha, D.S.; Steward, B.L. Shape and Size Analysis of Corn Plant Canopies for Plant Population and Spacing Sensing. *Appl. Eng. Agric.* **2005**, *21*, 295–303. [[CrossRef](#)]
31. Fan, J.; Zhou, J.; Wang, B.; de Leon, N.; Kaeppler, S.M.; Lima, D.C.; Zhang, Z. Estimation of Maize Yield and Flowering Time Using Multi-Temporal UAV-Based Hyperspectral Data. *Remote Sens.* **2022**, *14*, 3052. [[CrossRef](#)]
32. Brūmelis, G.; Dauškane, I.; Elferts, D.; Strode, L.; Krama, T.; Krams, I. Estimates of Tree Canopy Closure and Basal Area as Proxies for Tree Crown Volume at a Stand Scale. *Forests* **2020**, *11*, 1180. [[CrossRef](#)]
33. ten Harkel, J.; Bartholomeus, H.; Kooistra, L. Biomass and Crop Height Estimation of Different Crops Using UAV-Based LiDAR. *Remote Sens.* **2020**, *12*, 17. [[CrossRef](#)]

34. Maesano, M.; Khoury, S.; Nakhle, F.; Firrincieli, A.; Gay, A.; Tauro, F.; Harfouche, A. UAV-Based LiDAR for High-Throughput Determination of Plant Height and Above-ground Biomass of the Bioenergy Grass *Arundo Donax*. *Remote Sens.* **2020**, *12*, 3464. [\[CrossRef\]](#)
35. Li, X.; Liu, C.; Wang, Z.; Xie, X.; Li, D.; Xu, L. Airborne LiDAR: State-of-the-Art of System Design, Technology and Application. *Meas. Sci. Technol.* **2020**, *32*, 032002. [\[CrossRef\]](#)
36. Ravi, R.; Hasheminasab, S.M.; Zhou, T.; Masjedi, A.; Quijano, K.; Flatt, J.E.; Crawford, M.; Habib, A. UAV-Based Multi-Sensor Multi-Platform Integration for High Throughput Phenotyping. In Proceedings of the Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV, Baltimore, MD, USA, 15–16 April 2019. [\[CrossRef\]](#)
37. Wang, C.; Nie, S.; Xi, X.; Luo, S.; Sun, X. Estimating the Biomass of Maize with Hyperspectral and LiDAR Data. *Remote Sens.* **2017**, *9*, 11. [\[CrossRef\]](#)
38. Stuth, J.; Jama, A.; Tolleson, D. Direct and Indirect Means of Predicting Forage Quality through near Infrared Reflectance Spectroscopy. *Field Crops Res.* **2003**, *84*, 45–56. [\[CrossRef\]](#)
39. Shaver, R.D. *Evaluating Corn Silage Quality for Dairy Cattle*; University of Wisconsin—Madison Extension: Madison, WI, USA, 2007; pp. 1–11.
40. LaForest, L.; Hasheminasab, S.M.; Zhou, T.; Flatt, J.E.; Habib, A. New Strategies for Time Delay Estimation during System Calibration for UAV-Based GNSS/INS-Assisted Imaging Systems. *Remote Sens.* **2019**, *11*, 1811. [\[CrossRef\]](#)
41. Zhang, W.; Qi, J.; Wan, P.; Wang, H.; Xie, D.; Wang, X.; Yan, G. An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. *Remote Sens.* **2016**, *8*, 501. [\[CrossRef\]](#)
42. Woebbecke, D.M.; Meyer, G.E.; Von Bargaen, K.; Mortensen, D.A. Color Indices for Weed Identification under Various Soil, Residue, and Lighting Conditions. *Trans. Am. Soc. Agric. Eng.* **1995**, *38*, 259–269. [\[CrossRef\]](#)
43. Mardanisamani, S.; Maleki, F.; Kassani, S.H.; Rajapaksa, S.; Duddu, H.; Wang, M.; Shirtliffe, S.; Ryu, S.; Josuttes, A.; Zhang, T.; et al. Crop Lodging Prediction from UAV-Acquired Images of Wheat and Canola Using a DCNN Augmented with Handcrafted Texture Features. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 2657–2664. [\[CrossRef\]](#)
44. Kwak, G.H.; Park, N.W. Impact of Texture Information on Crop Classification with Machine Learning and UAV Images. *Appl. Sci.* **2019**, *9*, 643. [\[CrossRef\]](#)
45. Böhler, J.E.; Schaepman, M.E.; Kneubühler, M. Optimal Timing Assessment for Crop Separation Using Multispectral Unmanned Aerial Vehicle (UAV) Data and Textural Features. *Remote Sens.* **2019**, *11*, 1780. [\[CrossRef\]](#)
46. Duan, B.; Liu, Y.; Gong, Y.; Peng, Y.; Wu, X.; Zhu, R.; Fang, S. Remote Estimation of Rice LAI Based on Fourier Spectrum Texture from UAV Image. *Plant Methods* **2019**, *15*, 124. [\[CrossRef\]](#)
47. Van Der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Goullart, E.; Yu, T. Scikit-Image: Image Processing in Python. *PeerJ* **2014**, *2*, e453. [\[CrossRef\]](#)
48. Liu, N.; Li, L.; Li, H.; Liu, Z.; Lu, Y.; Shao, L. Selecting Maize Cultivars to Regulate Canopy Structure and Light Interception for High Yield. *Agron. J.* **2022**, *115*, 770–780. [\[CrossRef\]](#)
49. Song, Y.; Rui, Y.; Bedane, G.; Li, J. Morphological Characteristics of Maize Canopy Development as Affected by Increased Plant Density. *PLoS ONE* **2016**, *11*, e0154084. [\[CrossRef\]](#)
50. Petrović, N.; Moyà-Alcover, G.; Jaume-i-Capó, A.; González-Hidalgo, M. Sickle-Cell Disease Diagnosis Support Selecting the Most Appropriate Machine Learning Method: Towards a General and Interpretable Approach for Cell Morphology Analysis from Microscopy Images. *Comput. Biol. Med.* **2020**, *126*, 104027. [\[CrossRef\]](#)
51. Masjedi, A.; Crawford, M.M. Prediction of Sorghum Biomass Using Time Series UAV-Based Hyperspectral and Lidar Data. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Waikoloa, HI, USA, 26 September–2 October 2020; pp. 3912–3915. [\[CrossRef\]](#)
52. Jin, S.; Su, Y.; Zhang, Y.; Song, S.; Li, Q.; Liu, Z.; Ma, Q.; Ge, Y.; Liu, L.L.; Ding, Y.; et al. Exploring Seasonal and Circadian Rhythms in Structural Traits of Field Maize from Lidar Time Series. *Plant Phenomics* **2021**, *2021*, 9895241. [\[CrossRef\]](#)
53. Su, Y.; Wu, F.; Ao, Z.; Jin, S.; Qin, F.; Liu, B.; Pang, S.; Liu, L.; Guo, Q. Evaluating Maize Phenotype Dynamics under Drought Stress Using Terrestrial Lidar. *Plant Methods* **2019**, *15*, 11. [\[CrossRef\]](#)
54. Arnqvist, J.; Freier, J.; Dellwik, E. Robust Processing of Airborne Laser Scans to Plant Area Density Profiles. *Biogeosciences* **2020**, *17*, 5939–5952. [\[CrossRef\]](#)
55. Fonseca, A.E.; Westgate, M.E.; Grass, L.; Dornbos, D.L. Tassel Morphology as an Indicator of Potential Pollen Production in Maize. *Crop Manag.* **2003**, *2*, 1–15. [\[CrossRef\]](#)
56. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; NeurIPS: San Diego, CA, USA, 2017; pp. 5999–6009.
57. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)

58. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
59. Bernardo, R. Parental Selection, Number of Breeding Populations, and Size of Each Population in Inbred Development. *Theor. Appl. Genet.* **2003**, *107*, 1252–1256. [[CrossRef](#)]
60. Vigna, S. A Weighted Correlation Index for Rankings with Ties. In Proceedings of the WWW 2015: 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 1166–1176. [[CrossRef](#)]
61. Xie, C.; Yang, C. A Review on Plant High-Throughput Phenotyping Traits Using UAV-Based Sensors. *Comput. Electron. Agric.* **2020**, *178*, 105731. [[CrossRef](#)]
62. Lauer, J. Record When a Field Tassels to Predict Corn Silage Harvest Date. Available online: <https://ipcm.wisc.edu/blog/2013/07/record-when-a-field-tassels-to-predict-corn-silage-harvest-date/> (accessed on 1 June 2023).
63. Hoffman, P.C.; Lundberg, K.M.; Bauman, L.M.; Shaver, R.D. The Effect of Maturity on NDF Digestibility. *Focus Forage* **2003**, *5*, 1–3.
64. Hoffman, P.C.; Shaver, R.D.; Combs, D.K.; Undersander, D.J.; Bauman, L.M.; Seeger, T.K. Understanding NDF Digestibility of Forages. *Focus Forage* **2001**, *3*, 3–5.
65. Zhang, J.; Wang, X.; Zhang, H.; Sun, H.; Liu, X. Retrieval-Based Neural Source Code Summarization. In Proceedings of the ICSE'20: ACM/IEEE 42nd International Conference on Software Engineering, Melbourne, Australia, 27 June–19 July 2020; pp. 1385–1397. [[CrossRef](#)]
66. Masjedi, A.; Crawford, M.M.; Carpenter, N.R.; Tuinstra, M.R. Multi-Temporal Predictive Modelling of Sorghum Biomass Using Uav-Based Hyperspectral and Lidar Data. *Remote Sens.* **2020**, *12*, 3587. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.