agronomy

MDPI

Article

Leveraging Multi-Omics Data with Machine Learning to Predict Grain Yield in Small vs. Big Plot Wheat Trials

Jordan McBreen [1], Md Ali Babar [1,*], Diego Jarquin [1], Yiannis Ampatzidis [2], Naeem Khan [1], Sudip Kunwar [1], Janam Prabhat Acharya [1], Samuel Adewale [1] and Gina Brown-Guedira [3]

[1] Department of Agronomy, University of Florida, 3105 McCarty Hall B, Gainesville, FL 32608, USA
[2] Agricultural and Biological Engineering Department, Southwest Florida Research and Education Center, University of Florida, IFAS, 2685 SR 29 North, Immokalee, FL 34142, USA; i.ampatzidis@ufl.edu
[3] Plant Science Research, USDA-ARS SEA, Raleigh, NC 27695, USA
* Correspondence: mababar@ufl.edu

Abstract: Accurate grain yield (GY) prediction is essential in wheat breeding to enhance selection and accelerate breeding cycles. This study explored whether high-throughput phenotyping (HTP) data collected from small plot (SP) trials can effectively predict GY outcomes in later-stage big plot (BP) trials. Genomic (G) data were combined with hyperspectral (H) and multispectral + thermal (M) imaging across the 2022 and 2023 growing seasons at the Plant Science Research and Education Unit, Citra, Florida. A panel of 312 wheat genotypes was analyzed using GBLUP-based models, integrating G + H and G + M data from SP to predict BP yield. SP models demonstrated promising predictive ability, with G + H models achieving moderate within-year (0.43 to 0.51) and across-year (0.43) prediction accuracies, while G + M models reached 0.53 to 0.58 and 0.45, respectively. The Random Forest Regression (RFR) model produced an accuracy of 0.47 when M data from the 2022 SP, combined with G, was used to predict BP yield in 2023. Additionally, the top 25% specificity (coincide index) was evaluated, with models showing up to 47–51% within a year and 43–45% between years overlap in the highest predicted-yielding lines between SP and BP trials, further emphasizing the potential of SP data for early selection. These findings suggest that SP trials can provide meaningful predictions for BP yields, enabling earlier selection and faster breeding cycles.

Keywords: grain yield; wheat breeding; genomic prediction; high-throughput phenotyping; hyperspectral imaging; UAV phenotyping; small plot trials; machine learning

Academic Editors: Ana Fita, Jun Yan and Weilong Guo

Received: 18 March 2025
Revised: 15 May 2025
Accepted: 22 May 2025
Published: 28 May 2025

Citation: McBreen, J.; Babar, M.A.; Jarquin, D.; Ampatzidis, Y.; Khan, N.; Kunwar, S.; Acharya, J.P.; Adewale, S.; Brown-Guedira, G. Leveraging Multi-Omics Data with Machine Learning to Predict Grain Yield in Small vs. Big Plot Wheat Trials. *Agronomy* **2025**, *15*, 1315. https://doi.org/10.3390/agronomy15061315

1. Introduction

Wheat (*Triticum aestivum*) is a crucial global food crop, serving as a dietary staple for around 30% of the world's population [1,2]. Growing over 218 million hectares worldwide, it is crucial for food security in both developed and developing nations facing food scarcity. The rapidly growing global population and the escalating impacts of climate change are placing considerable pressure on wheat production systems. Current yields, averaging around 3300 kg ha^{-1}, are significantly threatened by rising temperatures and more unpredictable weather patterns [3,4]. In addition, wheat output must rise by 60% to accommodate an anticipated population of 10 billion by 2050, requiring an annual yield increase of no less than 1.6% [5,6]. Accomplishing this is not straightforward because it is impeded by abiotic stresses such as elevated temperatures, dryness, and unpredictable precipitation patterns, which significantly hinder wheat yield. Making accurate predictions of wheat yield is a primary goal for breeders, but can be complicated not only by abiotic

stress conditions but also by the variability existing between target environments and from plot to plot.

The classical breeding approach for yield improvement is still considered an informed "numbers game" where a large number of mid- and late-generation breeding lines are assessed in multi-location BP yield trials (also called yield trial plots). In the BP yield trials, GY per se is considered the main selection criterion. Along with environmental barriers, the BP trials are labor-intensive, costly, and time-consuming, often requiring several years and multiple locations to accurately assess yield stability. Nonetheless, traditional BP field trials are essential in the real-life scenario for identifying high-performing, suitable varieties for the target environments. To reduce these expenses and speed up the breeding cycle, breeders have long sought efficient tools that can provide early and reliable predictions of GY before initiating costly large-scale multi-location replicated trials. SP data, particularly from the mid-generation (F_5–F_7) stages of the breeding cycle, offers a promising solution.

Utilizing data collected on mid-generation breeding lines, especially from HTP platforms and single nucleotide polymorphism (SNP)-based molecular markers, could allow breeders to screen large numbers of genotypes quickly and cost-effectively before advancing only the most promising lines to later stages [7]. In doing so, not only can time and labor be reduced, but selection intensity can be increased as well. This can also pose a great benefit to breeders when seeds are limited during the multi-location field trials phase of their breeding cycles.

A primary problem that arises in breeding is the weak association between GYs assessed in SP and BP trials. Historically, the performance of wheat lines in the mid-generation breeding phases has not reliably predicted their yield in broader, more typical conditions. To tackle this issue, breeders use sophisticated phenotyping technologies like UAV-based spectral data, capable of capturing critical physiological parameters, including canopy reflectance and temperature indices [8]. When assessed in early yield trial stages (F_5–F_7 generation), these features may provide superior predictors of future yield potential in bigger plots, enabling breeders to make more educated choices at an earlier stage.

Developments in HTP have improved the capacity to estimate GY and other essential features in wheat, allowing researchers to take many types of data at all stages of the breeding cycle. UAV-based HTP systems, outfitted with multispectral, thermal, and hyperspectral sensors, provide breeders with an efficient and economical means to collect extensive phenotypic data [9–11]. UAV methods provide the acquisition of time-series data at critical phases of the wheat development cycle, enabling the early selection of advantageous genotypes [12]. Hinting at the potential for HTP data taken from lines earlier on in the breeding cycle to give information about the performance of key traits later in the process, namely GY. If this kind of data can accurately represent line performance at later stages, it could be used for making advancements of high-yielding lines and forego the unnecessary yield trials, or, adversely, for culling lines that will not end up performing well. Thus, allowing breeders to increase their selection intensity.

In HTP platforms, vegetation indices (VIs) like the normalized difference vegetation index (NDVI) and canopy temperature (CT) are recognized as dependable indicators of wheat biomass and yield potential when taken on sufficiently large plots. Multiple studies indicate that including HTP data (NDVI and CT) into genomic prediction models significantly improves yield prediction accuracy [13,14]. Additionally, canopy reflectance data, which offers insights into plant health and stress responses, has been shown to be useful to enhance the prediction capability of genomic prediction models [15]. These data sources have been shown to correlate with GY when measured in large plots, but their potential to predict future yield from SP trials remains underexplored.

Despite multispectral and thermal imaging providing insights into plant phenotyping and yield predictions, hyperspectral imaging (HSI) delivers a far more comprehensive look into the physiology of a given line. HSI acquires reflectance data over several small spectral bands, enabling the detection of tiny changes in plant health, stress responses, and physiological characteristics that may be missed by other imaging methods [16,17]. The extensive spectrum data allows HSI to assess physiological traits, including chlorophyll concentration, photosynthetic efficiency, and water stress, which are essential for predicting GY under adverse circumstances [18].

Studies have repeatedly shown that the integration of all hyperspectral bands acquired into genomic prediction models can significantly enhance prediction accuracy [19,20], compared to models that rely on vegetation indices calculated from a select number of bands. These models can effectively use hyperspectral data to predict future genotype performance, adeptly capturing intricate physiological responses to stress and environmental fluctuations [16,18]. Often surpassing conventional genomic models, or those that rely on phenomics data alone; HSI integration could help bridge the gap between SP traits and later-stage yield outcomes.

Machine learning (ML) methodologies, like Random Forest Regression (RFR) and gradient boosting regression (GBR), have shown efficacy in elucidating the intricate correlations between genomic and phenotypic data derived from BP experiments. These ML algorithms are adept at managing non-linear interactions between high-dimensional phenotypic data, such as hyperspectral indices, and SNP markers, rendering them optimal for predicting future performance based on early-stage data [21]. By integrating SP data with sophisticated ML methodologies, breeders might improve the precision of their predictions, thereby reducing the need for costly BP trials and facilitating the early selection of better lines.

Despite these advancements, there have been few studies that directly evaluate how data collected from small plot (SP) trials, especially in mid-generation lines, can dependably predict yield performances in larger, replicated big plot (BP) trials across multiple years. The ability to use SP-derived high-throughput phenotypic and genomic data for forward prediction would represent a significant step toward increasing the efficiency of early selection, particularly in stages where breeders are constrained by seed quantity or testing resources. This work aims to address that gap by assessing whether integrated multi-omic models trained on SP data can match or approximate the predictive power of BP-based models, ultimately guiding resource allocation and selection intensity at earlier stages of the breeding pipeline.

As touched on, one of the primary difficulties in wheat breeding is the expense and duration needed to perform yield trials on BP. These trials, executed throughout the advanced phases of the breeding cycle, are resource-demanding but needed for pinpointing lines that should be selected. Accurate yield estimates for BP trials using data from SP during the first phases of the yield trial process offer an opportunity for breeders to substantially save expenses, increase selection intensity, and expedite the selection process [8]. This research seeks to evaluate whether SP trials, in conjunction with HTP (such as NDVI, CT, and HSI) and genomics data, can produce predictions equal to those obtained from BP trials. By testing both within a single year and from year to year, we can see how temporal and environmental dynamics affect the predictions. It examines the predictive accuracy of GY in forward prediction scenarios, using data from one year to genotype performance in the following year. Through the integration of ML models to analyze the impact of various techniques on prediction accuracy, it also assesses whether models can capture complex interactions between genomic and phenotypic data from SP trials, improving their predictive capacity for BP performance.

2. Materials and Methods

2.1. Plant Genetic Materials and Experimental Design

Field tests were conducted throughout the growing seasons of 2021–22 (designated as 2022) and 2022–23 (designated as 2023) at the Plant Science Research and Education Unit (PSREU), University of Florida, Citra, FL, USA. A total of 312 facultative soft wheat advanced breeding lines, sourced from several wheat breeding programs around the southern United States, were assessed for this research (Supplementary Table S1). The breeding lines were established through the SunGrains™ cooperative breeding initiative, which includes contributions from the University of Arkansas, Clemson University, the University of Florida, the University of Georgia, Louisiana State University, North Carolina State University, and Texas A&M University. The SunGrains™ effort aims to generate wheat lines that are responsive to the diverse conditions of the participating institutions. The varied genotype panel included in this research facilitates representation across many settings in the southern U.S., making it particularly effective for assessing performance under diverse environmental conditions often seen in this area.

To predict model accuracy across varying plot sizes, the 312 genotypes were planted in two different plot types: smaller head row-sized plots (SP) and larger yield trial-sized plots (BP). SP included unreplicated head rows arranged in three rows, each measuring roughly 0.933 square meters (1.53 m × 0.61 m), while BP (7-row), measuring 5.58 square meters (3.96 m × 1.41 m), was machine-planted. Trials were planted in mid-November of 2021 and 2022, organized in an augmented block design with one replication and 15 sub-blocks, which included repeated check genotypes to facilitate valid comparisons. Both BP and SP trials had 390 plots with repeated checks (AGS 3015, AGS 2024, and AGS 2060) known to be widely adapted to the southeastern United States. The checks were replicated within each sub-block, making up around 20% of the total plots. Both plot sizes underwent analogous management practices, characterized by uniform applications of fertilizer, herbicide, fungicide, and irrigation. This research exposed the lines to terminal heat stress at the grain-filling phases, since Citra often encounters temperatures beyond 30 °C during the critical post-anthesis period. By growing the same genotypes in both SP and BP trials, the link between performance in the small head-row plots and that found in larger yield trials was assessed.

2.2. Trait Measurement and UAV-Derived HTP

Data were collected on days to heading (DTH), GY, and other UAV-based HTP metrics, including normalized difference vegetation index (NDVI), canopy temperature (CT), and hyperspectral imaging (HSI). The experimental location in Citra, FL, consistently received elevated ambient temperatures surpassing 30 °C throughout the grain-filling stage in both years, which can cause the genotypes to experience heat stress during the reproductive stages. DTH was documented as the duration in days from planting until 50% of the plants attained heading, using the Zadoks growth scale [22]. GY was assessed using a combine harvester for the BP and SP, with the harvested grain weight standardized to a moisture content of 13% and expressed in kg ha^{-1}. Yield values were recorded on a per-plot basis for over 300 individual genotypes across the two growing seasons, providing the dataset for model development and validation.

Multispectral NDVI and thermal CT data were acquired using a quadcopter UAV outfitted with the MicaSense Altum PT sensor, Shenzhen, China, which offers both multispectral and thermal imaging functionalities. UAV flights were executed twice throughout the growing season, with the first flight being conducted 5–7 days post-heading and the second flight happening two weeks later, aligning with critical growth phases indicative of GY [23]. The UAV operated at a height of 30 m and a velocity of 1.5 m per second

(m/s), achieving an 85% frontal overlap and a 70% lateral overlap between photographs to guarantee thorough coverage. Data were collected at solar noon and under clear weather to reduce the influence of cloud cover. The NDVI and CT data were processed using Pix4Dmapper to create orthomosaics, which were further analyzed in QGIS with the zonal statistics plugin to achieve plot-level averages for NDVI and CT.

Hyperspectral data were acquired with a hexacopter UAV equipped with a Resonon Pika L 2.4 hyperspectral camera (Resonon Inc., Bozeman, MT, USA). The UAV system operated at a typical height of 60 m and a speed of 1.5 m/s, with a front overlap of 85% and a side overlap of 70%, consistent with the multispectral and thermal flights. The Pika L camera acquires data within a spectral range of 380–1020 nm, segmented into 300 narrow bands. Two hyperspectral UAV flights were executed as well, with the timing remaining the same, where the first flight is about one week after heading, and the second flight two weeks thereafter. The data from both flights were averaged to account for temporal variance and minimize noise. The hyperspectral reflectance data were analyzed using Spectronon Pro software (version 3.4.11; Resonon Inc., Bozeman, MT, USA) for calibration and georectification, with regions of interest (RoIs) manually delineated for each plot. Figure 1 offers a visualization of the workflow for the UAV data collection process.

Both of the UAV-derived HTP datasets underwent calibration utilizing standard reflectance panels and were adjusted for radiometric consistency. For multispectral and thermal data, a calibrated reflectance panel was imaged at the beginning and end of each UAV flight to account for ambient light variability and sensor drift. Raw digital numbers were converted to surface reflectance using the Pix4Dmapper software (version 4.8.4; Pix4D S.A., Lausanne, Switzerland) that applies empirical line correction based on the reflectance panel values. The hyperspectral data were radiometrically corrected using Spectronon Pro software, incorporating dark current subtraction and flat-field correction. Dark current correction removed sensor noise from shutter-closed exposures, while flat-fielding addressed spatial variations in sensor sensitivity using lab-acquired calibration frames.

The spatial corrections conducted included image alignment, orthomosaic generation, and georeferencing. A structure-from-motion approach was used to stitch overlapping images into high-resolution orthomosaic images of the fields. Ground control points (GCPs) distributed across the field were collected using an RTK-GNSS receiver and manually linked in the image processing software to enhance spatial precision. Ortho-mosaics were georeferenced to the WGS84 coordinate system.

Multispectral, thermal, and hyperspectral datasets were all analyzed using zonal statistics to extract spectral fingerprints from each plot. These statistics were calculated in QGIS (version 3.28.3; QGIS Development Team, Open Source Geospatial Foundation Project) as well as the previously mentioned Spectronon Pro software by overlaying field plot shapefiles and extracting per-plot summaries across image layers. The high-dimensional HSI data includes hundreds of narrow spectral bands that serve as potential spectral indicators, which can be used as predictive covariates in multivariate genomic selection models. NDVI, CT, and reflectance data from HSI were ultimately combined with SNP marker data to build integrative models for predicting GY. All UAV-derived features were linked to ground truth yield measurements at the individual plot level using unique genotype and plot identifiers. The final dataset used in modeling included all entries, ensuring statistical robustness. Prior to modeling, all predictor variables were standardized (mean = 0, SD = 1). Each plot was georeferenced to a specific genotype using a master field layout file to ensure consistent tracking across all UAV flights and seasons. Spectral and thermal traits were averaged across replicate flights to account for day-to-day variation and reduce environmental noise. For modeling, all genotypic data entries were included, totaling 312 unique wheat lines

across two years. Figures showing only a limited number of points (e.g., n = 40) were designed for visual clarity and do not represent the actual validation dataset size.

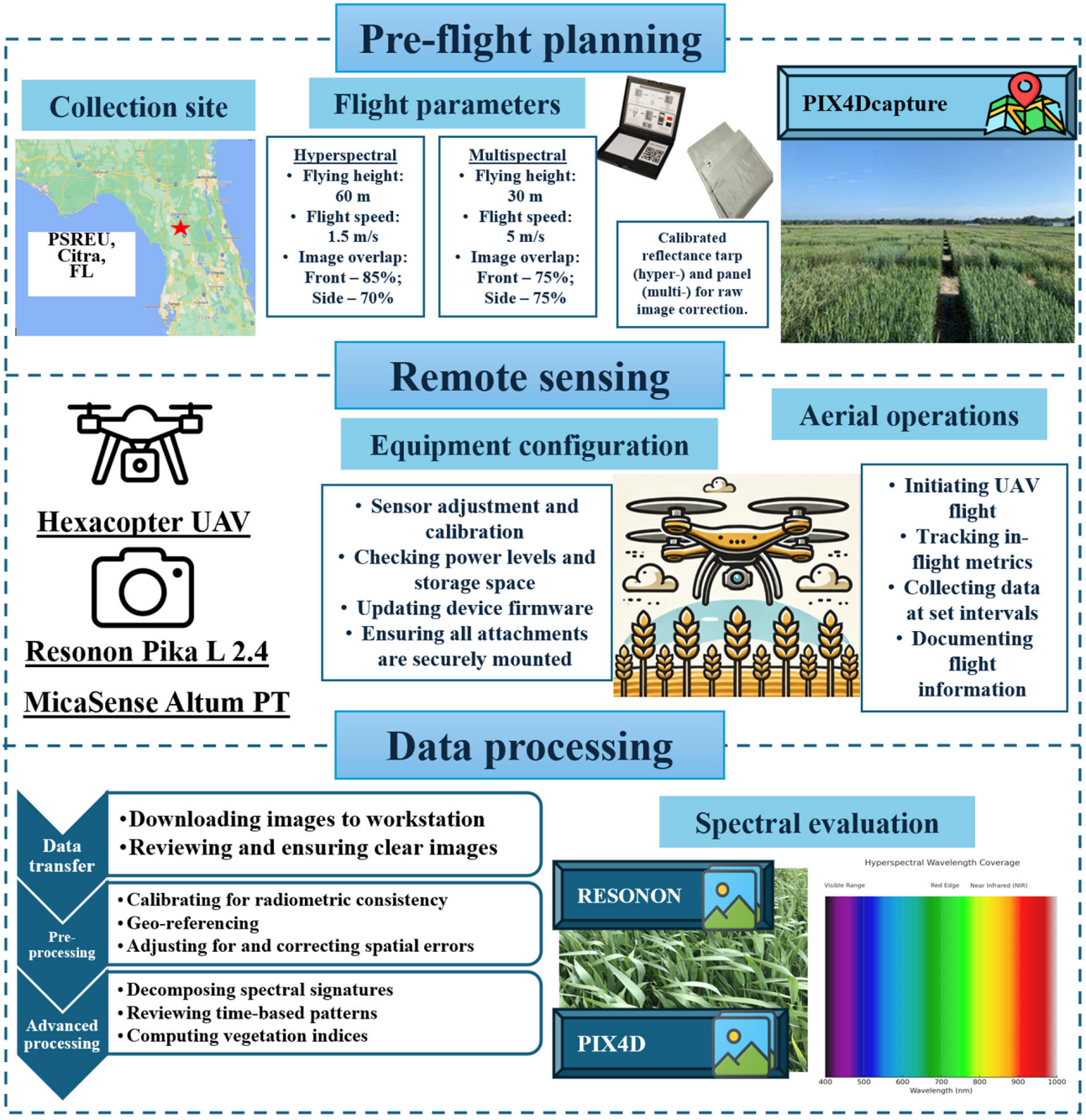

Figure 1. Workflow for UAV-based HTP (HSI, NDVI, and CT) data collection. Pre-flight planning includes setting flight parameters and using a calibrated reflectance tarp for image correction. Aerial operations involve tracking metrics and collecting data at set intervals. Data processing includes downloading images, radiometric calibration, spatial correction, and extracting vegetation indices.

2.3. Genotyping

Genetic characterization of wheat genotypes was conducted using a genotyping-by-sequencing (GBS)-based SNP markers. High-quality DNA was isolated from the leaf tissue of immature wheat seedlings using the sbeadex plant maxi kit (LGC Biosearch Technologies, Teddington, UK) on an oKtopure automated extraction equipment (LGC Genomics LLC, Teddington, UK). The GBS libraries were constructed via a two-enzyme digestion process with the restriction enzymes PstI and MspI. The fragmented DNA samples were ligated to

unique barcoded sequencing adapters for each sample. A total of 384 uniquely barcoded libraries were combined and sequenced using Illumina Novaseq 6000 SP flowcells, with a read length of 100 base pairs (Illumina Inc., San Diego, CA, USA).

Sequencing reads were aligned to the International Wheat Genome Consortium [24] RefSeqv1.0 reference genome using Burrows-Wheeler Aligner (BWA) version 0.7.12. SNP calling was performed with the TASSEL 5GBSv2 pipeline version 5.2.35. Markers exhibiting over 80% missing data, minor allele frequencies (MAFs) below 0.05, or heterozygosity above 10% were eliminated to maintain dataset quality. The filtering technique removed unreliable markers and enhanced the precision of future studies. The Beagle version 5.2 program was used for data imputation to resolve missing data issues. This approach utilizes linkage disequilibrium and k-nearest neighbor algorithms to predict absent genotypic values, hence enhancing the completeness of the genomic dataset. Following filtering and imputation, a final collection of 15,337 high-quality SNP markers was preserved, yielding a thorough genomic profile for each of the wheat genotypes.

2.4. Phenotypic Data Analysis

For each year, an ANOVA was carried out with the purpose of estimating the genotypic effect. The best linear unbiased estimates (BLUEs) were obtained for GY and the aerial-HTP derived spectral traits of NDVI, CT in degrees Celsius, and HSI data, separately for both the SP and BP trials. This is useful for removing biases from the fixed effects, ensuring that the effects of genotypes are properly estimated. For that, the R packages "lme4" version 1.1-7 and "emmeans" version 1.11.1 [25] were implemented. In the model used to extract the BLUEs, the genotypes were treated as fixed effects, while both environment and block were considered random.

The block and error terms were assumed to follow independent normal distributions. DTH was included as a covariate to adjust for any potential confounding effects due to differences in phenological development, as DTH is often linked to GY, NDVI, and CT. Along with the BLUEs, variance components were extracted to compute broad-sense heritability (H^2) by applying a model in which both genotype and block were used as random effects. From these variance components, H^2 was calculated for each trait within each environment. By estimating H^2, it is possible to quantify how much of the trait variation is attributable to genetic differences, providing information on how strongly the trait is controlled by genetics versus environmental factors. This helps assess the reliability of selecting certain traits under different environmental conditions. Separately analyzing each plot type (SP and BP) allows for an evaluation of genetic performance across environments and plot sizes.

2.5. Prediction Models

The research used five separate models to predict GY for wheat lines using several data types: genomic data (G), UAV-derived multispectral and thermal data (M), and UAV-acquired hyperspectral data (H). The BP and SP datasets were used separately for each model for yield prediction in BP trials. Genomic data were universally applicable across all lines, but M and H were used variably for each plot type. The models using BP-derived HTP data were designated B1 to B4, whilst those employing SP-derived HTP data were designated S1 to S4.

2.5.1. Genomic Data Model (G): B0

This model incorporated only G data (SNP markers) as predictors. SNP marker-based models are useful for identifying the genetic effects that contribute to yield, and this model

provides the baseline for comparison. The model was applied exclusively to BP yield prediction and did not incorporate any HTP data. The equation for model B0 is as follows:

$$Y_i = \mu + g_i + \varepsilon_i \quad (1)$$

where Y_i is the BLUE for GY of the ith genotype, μ is the general mean, g_i is the genomic effect of the i^{th} genotype, with the vector of genomic effects g following a multivariate normal distribution such that g~N(0, Gσ_g^2) and G $= \frac{XX'}{p}$ represents the genomic relationship matrix, and X is the standardized and centered (by columns) matrix of p SNPs, σ_g^2 is the corresponding variance component; and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ with σ_ε^2 is the error term variance.

2.5.2. Multispectral and Thermal Data Model (M): B1 and S1

This model integrated UAV-based NDVI and CT data (M) into the predictions for BP GY. These models captured phenotypic differences observable through aerial HTP data and allowed the assessment of how well spectral and temperature-based indices could predict yield. Model B1 utilized M data from BP, while S1 used the same data but taken from SP. The equation for this linear predictor was:

$$Y_i = \mu + M_i + \varepsilon_i \quad (2)$$

where M_i is the main effect of the NDVI and CT kernel for the i^{th} genotype, representing the phenotypic information derived from UAV-based sensors (Z). Where the joint distribution of the vector of phenomic effects is modeled as $M = \{M_i\} \sim N(0, \text{K}\sigma_K^2)$, such that $K = \frac{ZZ'}{m}$ represents a relationship matrix with Z as a matrix made up of the centered and standardized BLUE values of the m phenotypic traits and σ_P^2 denotes the corresponding variance component.

2.5.3. Genomic and Multispectral + Thermal Data Model (G + M): B2 and S2

To explore whether combining G with M data improves prediction accuracy, the G + M model incorporated both data types as predictors. Model B2 combined the data types when taken from BP trials, while S2 used SP data for yield prediction in BP. The equation for these models is as follows:

$$Y_i = \mu + g_i + M_i + \varepsilon_i \quad (3)$$

where the terms have been elaborated upon above, this model aimed to leverage both the genetic background and the environmental effects captured by HTP data to provide more precise predictions.

2.5.4. Hyperspectral Data Model (H): B3 and S3

These models were built using UAV-based hyperspectral data (H) to predict GY. Like before, model B3 utilized BP H data, while S3 utilized SP H data to predict BP yield. The equation is as follows:

$$Y_i = \mu + H_i + \varepsilon_i \quad (4)$$

where H_i represents the main effect of the hyperspectral data (S) for the i^{th} genotype, such that $H = \{H_i\} \sim N(0, \text{P}\sigma_P^2)$, where $P = \frac{SS'}{q}$ is the hyperspectral-derived relationship matrix with S as a matrix made up of the centered and standardized BLUE values of the q hyperspectral wavebands and σ_P^2 denotes the corresponding variance component.

2.5.5. Genomics and Hyperspectral Data Model (G + H): B4 and S4

Model G + H combined both data types to predict yield and assess whether the integration of both data types improves prediction accuracy. Model B4 used BP H data

alongside G, while S4 combined SP-derived H with G to predict BP yield. The equation for this model is as follows:

$$Y_i = \mu + g_i + H_i + \varepsilon_i \quad (5)$$

Through combining genetic and phenotypic data from hyperspectral imaging, this model aimed to improve the predictive power for yield under both BP and SP conditions.

2.5.6. Genomic, Phenomic, and Environmental Interaction Model (G × E, M × E, H × E): B2F, B4F, S2F, and S4F

To further enhance the prediction accuracy and account for the impact of environmental variability on yield, models incorporating environmental interactions (G × E, M × E, and H × E) were developed. These models integrate genomic (G), multispectral + thermal (M), and hyperspectral (H) data along with their interactions with the growing environment, defined here by year. The interaction terms allow for the assessment of how genotypes and phenotypic traits respond to differences in seasonal conditions. For forward predictions, these models were applied to predict BP yield across years. For instance: B2F incorporated G + M + G × E + M × E data from BP; B4F incorporated G + H + G × E + H × E data from BP; S2F incorporated G + M + G × E + M × E data from SP; S4F incorporated G + H + G × E + H × E data from SP. The equation for these models is as follows:

$$Y_i = \mu + g_i + P_i + (g_i \times E_i) + (P_i \times E_i) + \varepsilon_i \quad (6)$$

where g_i is the genomic effect for the i^{th} genotype; P_i is the phenotypic effect (M or H) for the i^{th} genotype; $g_i \times E_i$ is the genotype-by-environment interaction effect for the i^{th} genotype, capturing the variability of genetic performance across environmental conditions; $P_i \times E_i$ is the phenotypic-by-environment interaction effect for the i^{th} genotype, capturing the variability of phenotypic traits across environments; and ε_i is the residual error.

2.5.7. Machine Learning Models

For the across-year GY prediction, several machine learning (ML) techniques using genomic (G), hyperspectral (H), and multispectral plus thermal (M) data were tested. Support vector machine regression (SVMR) was implemented for its ability to handle high-dimensional datasets with relatively small sample sizes. Random Forest Regression (RFR) was evaluated for its ensemble-based architecture, which constructs multiple decision trees on random subsets of the data and averages their predictions to reduce variance and overfitting. Gradient boosting regression (GBR) was tested for its capacity to iteratively improve performance through additive model construction. Lastly, an Artificial Neural Network (ANN) model was employed to capture complex, non-linear interactions between predictors and grain yield.

Hyperparameter tuning for all ML models was performed using a grid search strategy within each training fold, with five-fold internal cross-validation to identify the optimal parameter set. For SVMR, the regularization parameter (C: 0.1, 1, and 10), kernel type (linear or radial basis function), and kernel coefficient (gamma: 'scale', 0.1, and 1) were evaluated. RFR models were tuned across the number of trees (n_estimators: 100, 200, and 500), the number of features considered at each split (max_features: 'sqrt' and 'log2'), and maximum tree depth (10, 20, or unrestricted). For GBR, tuning included learning rate (0.01, 0.05, and 0.1), number of estimators (100, 200, and 500), and tree depth (3, 5, and 10). ANN models were tested with various architectural and training configurations, including the number of hidden layers and neurons (e.g., 1 × 64, 2 × 64, and 2 × 128), activation functions (ReLU or tanh), batch size (32 and 64), learning rate (0.001 and 0.01), and optimizer (Adam). Final hyperparameter combinations for each model were selected

based on the configuration that minimized root mean square error (RMSE) on the internal validation set within each fold.

All multivariate data combinations (e.g., G + M, G + H, G + M + E, etc.) were structured by horizontally concatenating the standardized predictor matrices (genomic markers, vegetation indices, and/or hyperspectral bands) for each line into a single design matrix. This matrix was then used as input for each model. For ANN models, a feedforward fully connected architecture was used, where the final output layer contained a single neuron with a linear activation function for grain yield prediction. All predictors were standardized prior to model training, and missing values (if any) were imputed using mean imputation within each feature set.

2.6. Cross Validation

A CV2 strategy like the one outlined in Jarquin et al. [26] was used for within-year predictions. The population was divided into 10 clusters via discriminant analysis of principal components (DAPCs), which utilized year-specific SNP genotyping data to control for relatedness in the training and validation sets. This year-specific stratification ensured that clusters were unique to each year, avoiding any overlap of genotypes across years. The number of clusters was selected using the Bayesian Information Criterion (BIC), which identified the most parsimonious clustering configuration. PCA was separately used to visualize the population structure and confirm the presence of subgroup differentiation, but not for assigning clusters. This combined approach ensured that subpopulation structure was accounted for while preserving transparency in genetic diversity. The dataset was divided into five cross-validation subsets, with each subset serving as the validation set once, while the other four subsets constituted the training set. The procedure was carried out ten times, where 20% of the phenotypic information was concealed for validation inside each fold.

In the forward prediction scenario, a methodology was used to replicate authentic breeding circumstances, using data from previous years to predict performance in the following years. This predictive scenario evaluated the efficacy of models based on data from 2022 in predicting the yield of genotypes for the next year, 2023. This method offered insight into the models' capacity to generalize across annual variations. Conversely, since the same genotypes were grown in both years, we leveraged the 2023 data to predict the 2022 results as well. To evaluate model performance within each environment, the Pearson correlation (ρ) between predicted and observed values was calculated. The coincidence index (CI) was used to assess the effectiveness of different models in identifying top-performing genotypes in forward prediction scenarios. The CI quantifies the proportion of genotypes shared between the predicted and observed top 25% for grain yield, providing a practical measure of a model's utility in breeding programs where selecting superior genotypes is a key objective. To calculate the CI, the predicted rankings of genotypes were compared to their observed rankings for grain yield.

3. Results

3.1. Location and Weather

The trial was planted at PSREU in Citra, Florida, where the climatic circumstances are marked by elevated temperatures, especially during the crucial grain-filling phases of wheat development. Supplementary Figure S1 depicts the temperature trends during the growth seasons of 2021–2022 and 2022–2023. Within the two seasons, the average daily temperatures often reached 30 °C or above, with the highest temperatures observed from March to May. Higher temperatures during the reproductive phases can make the environment heat-stressed and offer a diverse environment for the panel to grow in.

Rainfall patterns seen in Supplementary Figure S2 exhibit variability throughout the two-year period. The 2022 season had increased precipitation occurrences in late December and mid-January, with significant maximum rainfall approaching 6 cm in a single day. In contrast, 2023 saw smaller, more frequent precipitation events, especially in February and March, with none above 3 cm per day. The variation in rainfall, coupled with elevated temperatures, signifies the many environmental stresses faced by the wheat lines, which might affect their phenological development and grain production performance.

3.2. Descriptive Statistics and Heritability

The descriptive statistics and H^2 for GY, NDVI, and CT over the 2022 and 2023 growing seasons and plot sizes are shown in Table 1. The data underscores the annual environmental variability and the genetic influences on the variables under consideration. For GY, SP had an average of 3854 ± 73 to 2473 ± 122 kg ha^{-1}, surpassing the 3355 ± 49 to 2332 ± 29 kg ha^{-1} seen in BP. The H^2 levels in BP ranged from 66 to 69%, while the range for SP was from 33 to 38%, suggesting a higher genetic influence on yield in the bigger plots. NDVI and CT exhibited similar tendencies, with BP demonstrating superior heritability (59 to 61% for NDVI and 49 to 55% for CT) relative to SP (51 to 54% for NDVI and 41 to 45% for CT), despite the mean trait values, in general, being higher for SP compared to BP with a single exception.

Table 1. Summary of the mean and SE and broad-sense heritability (H^2) of GY in kg ha^{-1}, NDVI, and CT in °C for BP and SP trials during the 2022 and 2023 growing seasons.

Year	Trait	Plot Size	Mean ± SE	H^2 (%)
2022	GY	BP	3355 ± 49	66
2022	NDVI	BP	0.78 ± 0.03	61
2022	CT	BP	35.7 ± 0.8	49
2022	GY	SP	3854 ± 73	38
2022	NDVI	SP	0.77 ± 0.04	54
2022	CT	SP	37.9 ± 0.13	41
2023	GY	BP	2332 ± 29	69
2023	NDVI	BP	0.66 ± 0.03	59
2023	CT	BP	31.2 ± 0.2	55
2023	GY	SP	2473 ± 122	33
2023	NDVI	SP	0.73 ± 0.03	51
2023	CT	SP	37.4 ± 0.7	45

BP, big plot; SP, small plot; NDVI, normalized difference vegetation index; CT, canopy temperature; GY, grain yield; SE, standard error.

H^2, obtained from the H data in the BP and SP, further illustrates the variability in genetic effects seen from 2022 to 2023 (Figure 2). In 2022, a notable increase in heritability was seen at the 700 nm wavelength, indicating heightened genetic impact within this spectral region. However, in 2023, this trend continued, and the H^2 values across hyperspectral wavelengths were typically reduced. The H^2 from the SP here shows similar trends to their BP counterparts for each year, but are generally lower.

The connection between GY, NDVI, and CT was evaluated by correlation analysis for each of the growing seasons (Figure 3). For 2022, there was a moderate positive correlation found between NDVI and GY in the BP (ρ = 0.56 ***), which may be suggestive that higher vegetation index values, which are reflective of greater biomass or canopy greenness, are associated with higher yields under the environmental conditions within this year. However, for CT, the correlation with GY was negative, indicative of elevated CT corresponding to reduced GY values. These patterns in the correlation between traits were similar in the SP, with GY positively correlated with NDVI (ρ = 0.51 ***), while the

correlation between GY and CT was negative (ρ = −0.61 ***). The trend persisted in the next year, reinforcing the association between increased canopy greenness and yield potential, though the correlation between GY and CT exhibited greater variability in 2023. The persistent positive association between GY and NDVI during both years indicates that NDVI serves as a dependable predictor of grain yield potential. The negative connection between GY and CT emphasizes the adverse effect of elevated temperatures on yield, highlighting the significance of canopy cooling as a trait for enhancing varieties adapted to the hot, humid environment. The correlation between BP and SP's GY was generally low to very low.

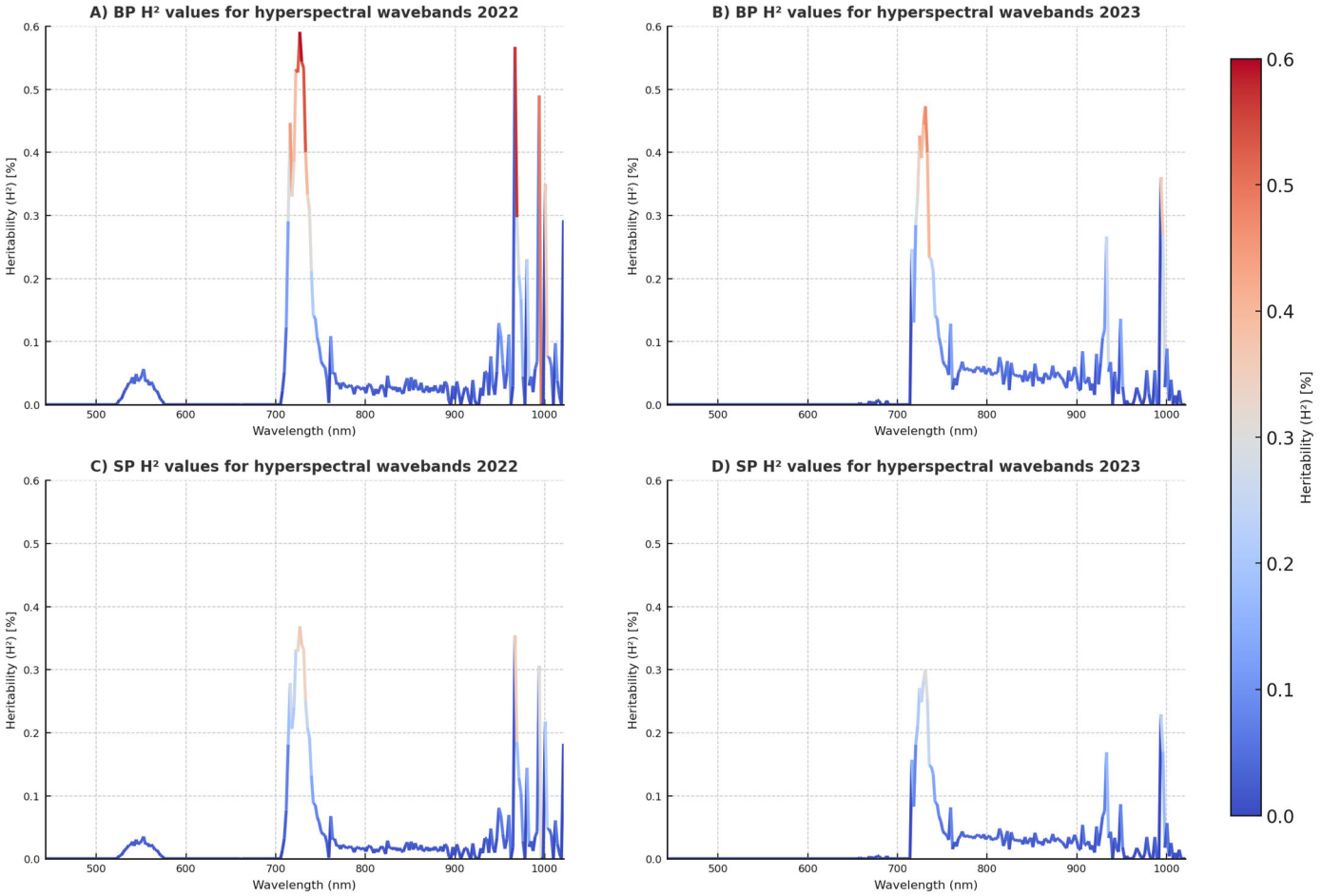

Figure 2. The broad sense heritability (H^2) values for hyperspectral wavebands in (**A**) 2022 BP, (**B**) 2023 BP, (**C**) 2022 SP, and (**D**) 2023 SP. SP, small plots; BP, big plots.

3.3. Model Evaluation

3.3.1. Model Stratification Results

The DAPC analysis was carried out to examine and control the effects of population structure in the wheat panel used. Figure 4 shows the PCA plot to visualize the genetic diversity captured within the population. The clusters in the illustration are distinguished by color, with each hue denoting a distinct subpopulation determined by the model. The first two principal components (PC1 and PC2) account for 5.2% and 4.2% of the variance, respectively, highlighting the genetic diversity within the population. Scattering of genotypes shows more genetic diversity and highlights the relevance of controlling for structure in the analysis. To complement this analysis, a phylogenetic tree was constructed based on genome-wide SNP data to visualize genetic relationships among genotypes (Supplementary Figure S3), further supporting the observed population structure.

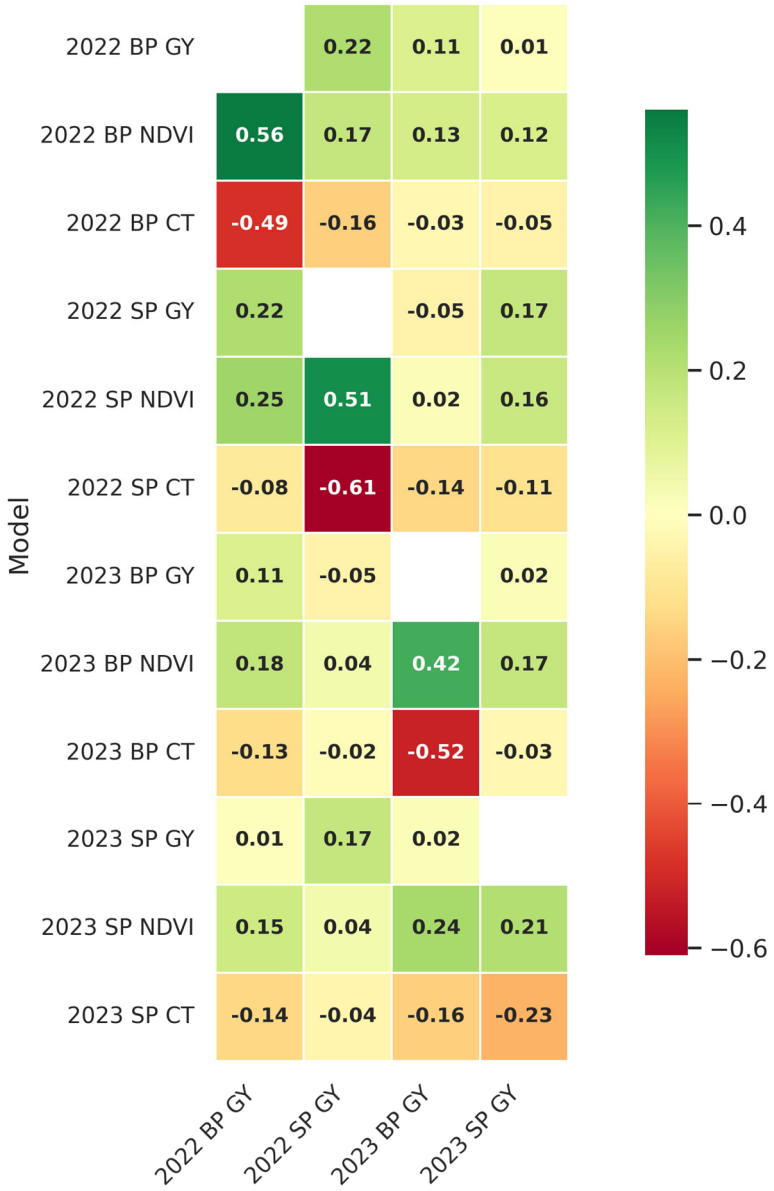

Figure 3. Heatmap of Pearson correlation coefficients between GY, NDVI, and CT across SP and BP trials for the 2022 and 2023 growing seasons. GY, grain yield; NDVI, normalized difference vegetation index; CT, canopy temperature; BP, big plot; and SP, small plot. Correlation coefficient values above 0.43, 0.32, and 0.20 are significant at the 0.001, 0.01, and 0.05 probability levels, respectively.

3.3.2. Within-Year Model Performance

In Table 2, the prediction accuracies of several models predicting GY in BP over the 2022 and 2023 growing seasons are displayed. The table shows the prediction accuracy of each model using the Pearson correlation (ρ) between the observed GY and its predicted values. The evaluated models consist of some inclusion or combination of G, M, and H data types. These models (G, M, H, and G + M or G + H) are assessed to determine the advantages of using complementary datasets. The plot data column denotes whether the high-throughput data were derived from BP or SP, and the data type describes the data included. Model numbers B0, B1–B4, and S1–S4 denote which model is used and from which data source, and the description column explains what each model is performing. Each model's performance is assessed using a 10-fold cross-validation method, with an 80/20 training-testing split, according to the CV2 cross-validation technique.

Figure 4. Principal Component Analysis (PCA) of genotypic variation among 2023 wheat lines based on genome-wide SNP markers. Each point represents a genotype and is color-coded by one of 10 groups determined via discriminant analysis of principal components (DAPCs). The PCA is shown here solely for visualization of genetic structure, while subgroup assignment was carried out independently using DAPCs. The number of clusters (k = 10) was selected based on the lowest Bayesian Information Criterion (BIC) value obtained from the clustering algorithm.

Table 2 shows how the models containing G and HTP data (G + M or G + H) often surpass models that depend only on a singular data source. For 2022, the G + M model (B2) had an accuracy of 0.57, and the G + H model (B4) narrowly surpassed it with a correlation of 0.61. For the following year, model B2 achieved 0.65 correlation, while B4 did marginally better, achieving an accuracy of 0.67, demonstrating how both genomic data and known correlated HTP data can complement one another.

Models relying on SP data exhibited marginally reduced prediction accuracies in comparison to the BP data models. The G + M model (S2) applied to BP data in 2022 attains a correlation of 0.53, while the G + H model (S4) gives a correlation of 0.43. In 2023, the G + M model (S2) achieves a correlation of 0.58, whereas the G + H model (S4) is 0.51. The differences found in the accuracy between BP and SP models may be due to increased variability that is found and probable measurement noise inherent in the SP data. The G + M and G + H models regularly surpass the simpler genomic-only or HTP-only models. The superior performance of these integrated models justifies their use in forward prediction situations since they provide a more holistic approach to predicting GY.

Table 2. Correlation (ρ) and SE between observed and predicted yield values of nine models predicting GY of BP in kg ha^{-1} within the 2022 and 2023 growing seasons in Citra, FL.

Year	Plot Data	Data Type	Model	Data Predicted	Model #	Corr (ρ)	SE	Model Description
2022	-	Genomic (G)	BP: G	BP yield	B0	0.33	0.05	Genomic marker data used to predict big plot yield within the year
2022	BP	NDVI and CT	BP: M	BP yield	B1	0.28	0.06	NDVI and CT data taken from BP used to predict BP yield within the year
2022	BP	G + NDVI and CT	BP: G + M	BP yield	B2	0.57	0.04	Genomic marker + NDVI and CT data taken from BP used to predict BP yield within the year
2022	BP	Hyperspectral (H)	BP: H	BP yield	B3	0.34	0.05	Hyperspectral data taken from BP used to predict BP yield within the year
2022	BP	G + H	BP: G + M	BP yield	B4	0.61	0.03	Genomic marker + hyperspectral data taken from BP used to predict BP yield within the year
2022	SP	NDVI and CT	SP: M	BP yield	S1	0.19	0.04	NDVI and CT data taken from SP used to predict BP yield within the year
2022	SP	G + NDVI and CT	SP: G + M	BP yield	S2	0.53	0.02	Genomic marker + NDVI and CT data taken from SP used to predict BP yield within the year
2022	SP	Hyperspectral (H)	SP: H	BP yield	S3	0.21	0.03	Hyperspectral data taken from SP used to predict BP yield within the year
2022	SP	G + H	SP: G + H	BP yield	S4	0.43	0.05	Genomic marker + hyperspectral data taken from SP used to predict BP yield within the year
2023	-	Genomic (G)	BP: G	BP yield	B0	0.36	0.04	Genomic marker data used to predict big plot yield within a year
2023	BP	NDVI and CT	BP: M	BP yield	B1	0.32	0.06	NDVI and CT data taken from BP used to predict BP yield within the year
2023	BP	G + NDVI and CT	BP: G + M	BP yield	B2	0.65	0.05	Genomic marker + NDVI and CT data taken from BP used to predict BP yield within the year
2023	BP	Hyperspectral (H)	BP: H	BP yield	B3	0.44	0.06	Hyperspectral data taken from BP used to predict BP yield within the year

Table 2. *Cont.*

Year	Plot Data	Data Type	Model	Data Predicted	Model #	Corr (ρ)	SE	Model Description
2023	BP	G + H	BP: G + H	BP yield	B4	0.67	0.05	Genomic marker + hyperspectral data taken from BP used to predict BP yield within the year
2023	SP	NDVI and CT	SP: M	BP yield	S1	0.27	0.03	NDVI and CT data taken from SP used to predict BP yield within the year
2023	SP	G + NDVI and CT	SP: G + M	BP yield	S2	0.58	0.05	Genomic marker + NDVI and CT data taken from SP used to predict BP yield within the year
2023	SP	Hyperspectral (H)	SP: H	BP yield	S3	0.24	0.05	Hyperspectral data taken from SP used to predict BP yield within the year
2023	SP	G + H	SP: G + H	BP yield	S4	0.51	0.03	Genomic marker + hyperspectral data taken from SP used to predict BP yield within the year

BP, big plots; SP, small plots; GY, grain yield.

3.3.3. Across-Year Forward Prediction Model Performance

In the forward prediction scheme that used data from one year (2022) to predict GY for the next year (2023), the G + M and G + H models were evaluated due to their within-year superior performance. Table 3 presents the predictive accuracies of different models using 2022 data to predict the 2023 BP yield. The model number is the same as in Table 2, aside for the inclusion of F at the end to denote forward/across-year prediction. For model B2F, G + M data taken from BP in 2022, plus environment interactions (G×E and M×E), were used to predict BP yield in 2023. In model B4F, G + H taken from BP in 2022, plus environment interactions (G×E and H×E), were used to predict BP yield in 2023. For S2F (G + M) and S4F (G + H), data taken from SP in 2022, plus environment interactions, were used to predict BP yield in 2023. The B2F attains a correlation of 0.47, but the B4F model yields a somewhat higher accuracy of 0.51. Indicating that the amalgamation of genetic data with HTP data enhances predictions over the years, with H increasing forward prediction accuracy relative to M. For SP, the associated forward prediction models (S2F and S4F) demonstrate slightly reduced accuracies, with S2F attaining 0.45 and S4F producing 0.43. For the reverse prediction, where 2023 data were used to predict 2022 BP yield, slightly lower predictive accuracies were observed. Model B2F (G + M from BP in 2023) attained a correlation of 0.42, while model B4F (G + H from BP in 2023) achieved 0.44. Similarly, the SP-based forward prediction models (S2F and S4F) using 2023 data yielded correlations of 0.41 and 0.39, respectively.

Table 3. The predictive accuracies between actual and predicted values of the two top-performing models from each plot data source: BP and SP when predicting GY in kg ha^{-1}.

Method	Plot Data Source	Data Type	Model	Data Predicted	Model #	Correlation (ρ)
FwdPred	2022BP	G + NDVI and CT	BP: G + M + G×E + M×E	2023BP yield	B2F	0.47
FwdPred	2022BP	G + H	BP: G + H + G×E + H×E	2023BP yield	B4F	0.51
FwdPred	2022SP	G + NDVI and CT	SP: G + M + G×E + M×E	2023BP yield	S2F	0.45
FwdPred	2022SP	G + H	SP: G + H + G×E + H×E	2023BP yield	S4F	0.43
FwdPred	2023BP	G + NDVI and CT	BP: G + M + G×E + M×E	2022BP yield	B2F	0.42
FwdPred	2023BP	G + H	BP: G + H + G×E + H×E	2022BP yield	B4F	0.44
FwdPred	2023SP	G + NDVI and CT	SP: G + M + G×E + M×E	2022BP yield	S2F	0.41
FwdPred	2023SP	G + H	SP: G + H + G×E + H×E	2022BP yield	S4F	0.39

The models in this table contain G, M (NDVI + CT), and H. BP, big plots; SP, small plots; G, genomic data; NDVI, normalized difference vegetation index; CT, canopy temperature; H, hyperspectral data.

We selected the four multi-kernel models (B2, B4, S2, and S4) based on their performance (Tables 2 and 3) for further investigation to improve prediction accuracies by using different ML models. The ML models include RFR, SVMR, GBR, and ANN, where the solid blue trendline represents the line of best fit, indicating the general trend of the predictions, while the shaded area around the trendline represents the 95% confidence interval, providing an estimate of uncertainty around the predictions. The combined panel shown in Figure 5 illustrates the actual versus predicted values across these models for each selected data configuration. Panel A shows predictions from the B2F model (G + NDVI +

CT from BP), where SVMR achieved the highest accuracy with a correlation of 0.50, while ANN yielded the lowest (ρ = 0.35). While the trendlines indicate the models generally captured the relationship between actual and predicted yields, some variability remains.

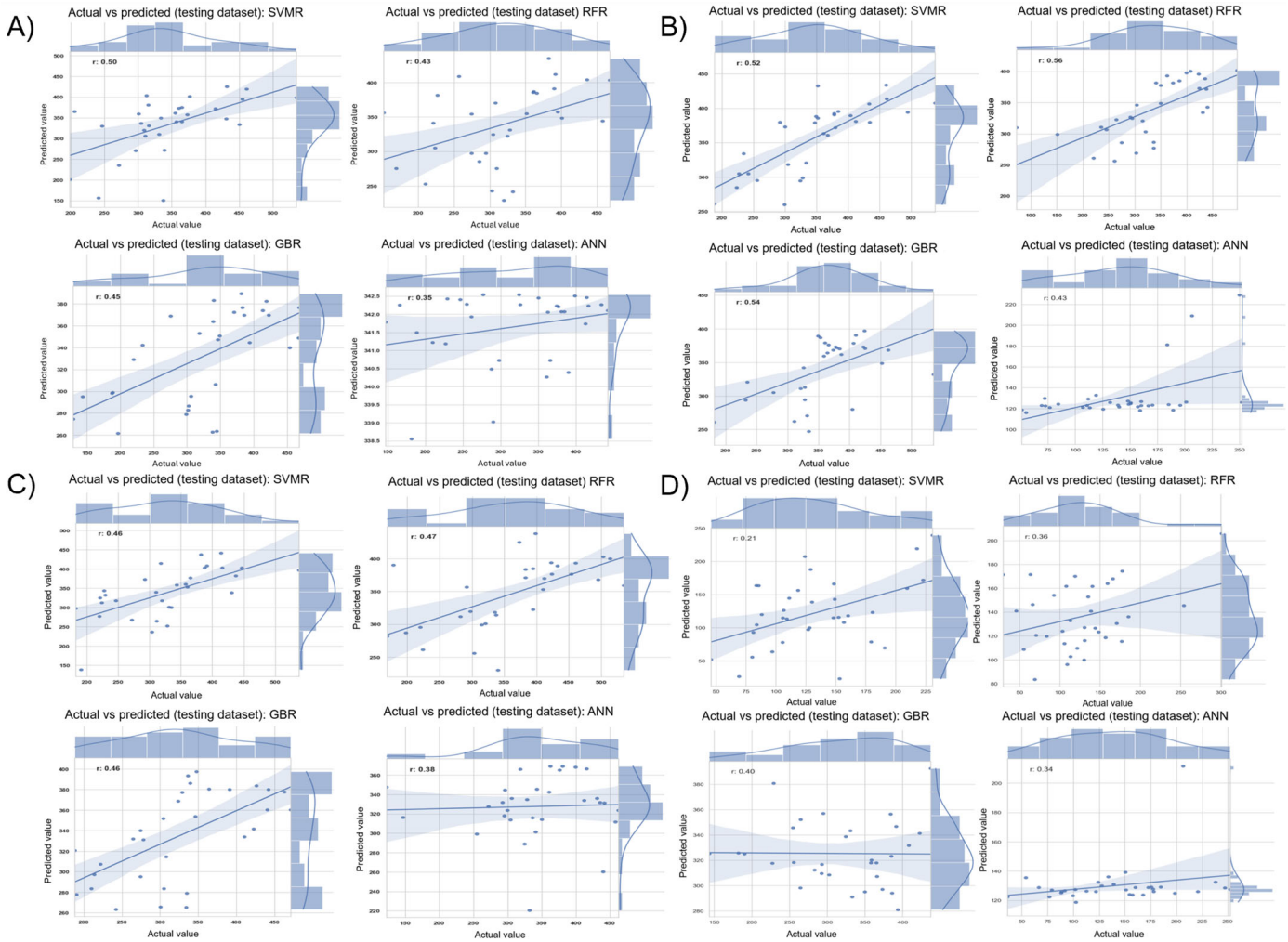

Figure 5. Each panel (**A**–**D**) contains results from four ML models: support vector machine regression (SVMR), random forest regression (RFR), gradient boosting regression (GBR), and artificial neural network (ANN). (**A**) G + NDVI + CT from BP in 2022; (**B**) G + H from BP in 2022; (**C**) G + NDVI + CT from SP in 2022; (**D**) G + H from SP in 2022. The blue trendline represents model fit with 95% confidence interval shading. Each subpanel shows a representative subset of 40 genotypes for visualization clarity. ML = machine learning, G = genomic, NDVI = normalized difference vegetation index, CT = canopy temperature, H = hyperspectral data, BP = big plots, SP = small plots.

Panel B evaluates the B4F model (G + H from BP), where RFR attained the highest correlation (r = 0.56), with GBR and SVMR also performing well (r = 0.54 and 0.52, respectively). ANN again had the lowest accuracy (r = 0.43). In Panel C, which evaluated S2F model was evaluated and found that SVMR, RFR, and GBR achieve similar correlations of 0.46, 0.47, and 0.46, respectively. This demonstrates the relatively stable performance of the ML models across data sources but also underscores the consistent underperformance of ANN. Panel D displays the S4F model (Figure 5), where the RFR model has superior performance with a correlation of 0.45, followed by the SVMR and GBR models with correlations of 0.32 and 0.44, respectively. ANN persistently underperforms in these contexts, with a correlation of 0.36.

An investigation was performed to compare the accuracy of SP-based models in predicting the 25% highest actual yielding lines from the BP trials. Table 4 presents the

predictive ability of the SP model (S2) by summarizing the percentage of top 25% yielding lines that were shared by SP predicted and BP actual GY data within and between years. This was considered the model specificity. We used the S2 model as it showed the highest predictive accuracy within and between years in our previous analysis. In 2022, the S2 model accurately recognized 51% of the top lines, while in 2023, it identified 48%. In making forward predictions on 2023 BP yield with 2022 SP data, the traditional S2F model demonstrated a specificity of 43%, whereas the RFR model using this same data enhanced this to 45%.

Table 4. Specificity of SP-based models in predicting the top 25% highest-yielding wheat lines from BP trials within different years and forward prediction scenarios.

Model	Training Plot Size	Predicted Plot Size	Accuracy	Top 25% Specificity
S2 (G + M)	2022 SP	2022 BP	0.53	51%
S2 (G + M)	2023 SP	2023 BP	0.58	48%
S2F (G + M)	2022 SP	2023 BP	0.45	43%
S2F (G + M) (RFR)	2022 SP	2023 BP	0.47	45%

BP, big plots; SP, small plots; G, genomic data; M, NDVI + CT data; RFR, Random Forest Regression.

The number of genotypes that overlapped between the top 25% of predicted highest yielding lines for the BP and SP-derived models, both within each year and in the forward prediction scenarios, was compared (Table 5). For BP and SP, we used B4 and S2 models, respectively, as those showed the highest prediction accuracies in our previous analysis. For example, the top 25% highest predicted yielding lines for the 2022 year using the BP-derived data were compared to the top 25% highest predicted yielding lines for the 2022 year using the SP-derived data, and the percent overlap between the line rankings was recorded. The overlap between the top 25% of lines predicted by BP and SP models within each year is relatively high, with a 67% overlap in 2022 and 71% in 2023. The forward prediction scenarios, where models from one year (2022) are used to predict the top 25% yielding lines in the following year (2023), show a lower degree of overlap, with a 57% overlap between B4F and S2F and 49% when using RFR.

Table 5. Overlap in the top 25% of predicted highest-yielding wheat lines between BP and SP models across different growing seasons and prediction scenarios.

Models Compared		Top 25% Overlap
2022 B4 (G + H)	2022 S2 (G + M)	67%
2023 B4 (G + H)	2023 S2 (G + M)	71%
B4F (G + H)	S2F (G + M)	57%
B4F (G + H) (RFR)	S2F (G + M) (RFR)	49%

The table shows the percentage overlap in top-performing wheat lines predicted by models using G + H for BP and G + M for SP. The overlap is calculated within the same year and in forward prediction scenarios, with RFR models used in some cases to improve predictions. This comparison highlights the consistency of SP models in identifying top-performing lines, even when used to predict across different years and plot sizes. BP, big plots; SP, small plots; G, genomic data; M, NDVI + CT data; H, hyperspectral data; RFR, Random Forest Regression.

4. Discussion

One of the primary trends emerging from our findings was the difference found in predictive power between the SP- and BP-based models. Expectedly, the BP models, which benefit from larger plot sizes and greater environmental representation, generally showed superior GY prediction accuracies. With the increased plot size of GY trials, the advantage of capturing more environmental heterogeneity leads to more reliability in data collection, as observed in previous research [27]. This sense of heightened environmental variability allows the BP-based modeling approaches to be better at generalizing across

different locations and the conditions found therein. This is something vital for predicting yield stability across multiple years. However, the findings of our research imply that SP models have the potential to narrow this performance gap, especially when said models are enriched with the incorporation of HTP data. Although the SP trials by nature contain less environment variability, including UAV-based phenotyping sources like NDVI, CT, and HSI, as was performed in this research, allows for the SP models to capture a more exhaustive physiological profile of the genotypes being tested. Such added phenotypic depth can make up for the smaller plot size and make SP trials more viable for mid-stage selection [28]. Our results suggest that although BP-based models typically achieve higher prediction accuracy, the SP-based models, when combined with HTP data, particularly NDVI and CT, performed reasonably well in comparison. Conveying the idea that even with the limitations inherent to the SP trials, when coupled with advanced phenotyping tools, they can still provide meaningful insights into GY selection.

Integrating HTP data sources into our approaches is critical for improving SP-based model performance, and though HSI is often seen as a powerful tool, in our current study, it did not manage to outperform more traditional methods like NDVI and CT. HSI offers high-dimensional phenotypic data by capturing reflectance across hundreds of narrow spectral bands, allowing it to detect subtle physiological traits that are closely linked to yield potential and are difficult to measure using simpler indices like NDVI and CT. However, HSI comes with significant drawbacks, including higher costs, greater technical complexity, and the need for more extensive data processing. In our case, despite its theoretical advantages, the SP models that relied on NDVI + CT outperformed those using HSI, suggesting that, under certain conditions, NDVI and CT are more practical and equally or more effective in capturing the key traits necessary for predicting grain yield from SP data [18,29].

The predictive accuracies observed in this study align well with findings from other wheat and small grain breeding studies that integrate genomic and HTP data. For example, Kaur et al. [30], and Krause et al. [19] demonstrated that combining genomic data with spectral indices or hyperspectral imaging can enhance grain yield predictions, reporting correlations ranging from 0.4 to 0.6 in stress environments. Similarly, Rutkoski et al. (2016) [31] achieved accuracies between 0.4 and 0.7 using multi-trait genomic prediction models integrated with hyperspectral data. Our results, such as a forward prediction correlation of 0.51 for G + H models, are consistent with these findings, especially considering the use of data from a single environment. Further endeavors can incorporate multi-location trials could validate these findings further and explore their generalizability across diverse environments.

The physiological indicators of plant health, especially when identified in early-stage studies, provide a crucial insight into the genotype's prospective efficacy in BP trials [32]. Despite HSI data by nature being more information-dense, findings indicate that NDVI and CT, while effective for evaluating overall plant biomass and temperature, marginally surpassed HSI data integration in forward prediction scenarios using the SP-derived HTP data. The G + M model for SP forward prediction had a slightly superior accuracy ($\rho = 0.45$) compared to the G + H model ($\rho = 0.43$). This indicates that, in SP trials, NDVI and CT may just as or even more successfully capture essential features for yield prediction as does hyperspectral data in cross-year predictions. For the BP trials, HSI exhibited more predictive capability ($\rho = 0.51$) than NDVI and CT ($\rho = 0.47$), suggestive of HSI's proficiency in identifying nuanced physiological responses essential for yield predicting in more expansive, heterogeneous contexts.

ML models showed variability in evaluating the prediction capability of the combined genomic and HTP data [33]. The research examined different ML models, with RFR

and GBR regularly identified as the top performers. These tree-based models proficiently capture the non-linear interactions among genetic, phenotypic, and environmental data, which are essential for managing high-dimensional information produced by UAV-based HTP systems. Our findings indicated that both RFR and GBR managed noisy data more effectively than SVMR and ANN, which encountered difficulties in situations characterized by significant environmental fluctuation, particularly with SP data. The enhanced effectiveness of RFR and GBR is likely attributable to their capacity to consolidate weaker predictors and decrease variation, which is essential in settings marked by significant volatility. Conversely, ANN models, which are prone to overfitting short datasets, were less adept at elucidating the intricate connections within our data, a phenomenon also noted in related yield prediction research [34]. SVMR, while competitive, underperformed in our setting, perhaps owing to its need for bigger, more structured datasets to surpass tree-based models. While hyperparameter refinement was carried out to minimize overfitting during model training, the validation of these models was limited to performance in the same location (Citra, FL) across different years. Future studies leveraging multi-environment trial data could help confirm the broader applicability of these models by assessing their performance across diverse locations. The findings highlight the efficacy of ensemble tree-based methods for genomic prediction, especially in breeding programs functioning within intricate and variable contexts.

The specificity findings indicated that SP models integrating M and G data attained a maximum overlap of 51% with the highest-yielding lines that were observed in the BP trials. Showing that, despite reduced plot sizes and decreased environmental representation, SP models trained on G + M data may still effectively capture a significant amount of the top-performing lines. For forward prediction, using 2022 SP data to forecast 2023 BP performance, the specificity remained relatively high, with 43% to 45% of the top 25% lines being consistently recognized by the SP models for the BP yields. SP models, even when used for multi-year predictions, provide significant insights into high-yielding lines, although with a little decrease in predictive accuracy relative to within-year projections.

SP-derived models successfully identified a significant proportion of the high-yielding wheat lines recognized by BP-derived models, especially within the same growing season. The overlap in the top 25% of anticipated highest-yielding lines between SP and BP models demonstrates the potential of SP data to contribute to early selection decisions. For forward predictions, this overlap decreased, likely due to increased temporal variability influencing model predictions based on data from BP or SP trials. Nevertheless, SP models consistently provided meaningful insights for predicting GY performance in future years.

While achieving prediction accuracies of around 0.5 may seem modest, it is reasonable within the context of early-generation wheat breeding, where limited replication, small plot sizes, and consistent environmental variability appear to constrain model performance. In early-stage selection, where field trials are often unreplicated and constrained by limited seed or land, models with predictive correlations above 0.4 can provide enough signal to support enrichment of superior lines and culling of low performers. This is particularly relevant when HTP and genomic data are combined, as these approaches allow breeders to identify candidates worth advancing even before resource-intensive large plot trials are initiated.

In terms of predictor contribution, NDVI consistently emerged as a key variable across models, reflecting its well-established utility as a proxy for canopy vigor, photosynthetic efficiency, and overall biomass—factors closely tied to grain yield. In this study, we prioritized NDVI as the primary vegetation index due to its longstanding utility in yield prediction, biological interpretability, and consistent performance across environments. NDVI served as a reliable and widely accepted proxy for canopy vigor and biomass in

our modeling framework. In practical terms, even moderate prediction accuracy allows breeders to eliminate a significant portion of underperforming lines, increasing selection intensity and reducing field costs without sacrificing long-term genetic gain. Prior studies have shown that genomic selection can deliver meaningful gains at similar accuracy levels [35] and that multi-trait genomic models can further improve the prediction of complex traits like grain yield [36]. While expanding the set of vegetation indices may marginally improve model performance, our findings indicate that NDVI alone captures the essential physiological variation relevant to yield under the conditions tested.

In terms of variable influence, vegetation indices like NDVI and CT consistently emerged as strong contributors in models using M data, aligning with their known associations with canopy vigor and thermal stress under heat conditions. In the hyperspectral models, reflectance features within the red-edge (~700–740 nm) and NIR (~750–900 nm) bands appeared to drive predictive performance, likely due to their linkage to chlorophyll content, biomass, and water status. Ensemble models such as RFR and GBR naturally provided variable importance rankings, from which these trends were inferred. The improvement in accuracy observed when combining genomic and phenomic inputs also suggests that these data types capture complementary biological signals relevant to yield performance across years.

The practical value of this framework lies in its ability to support early-stage selection decisions in wheat breeding using SP data, thereby reducing the resource burden associated with large-scale yield testing. By leveraging predictive models built on SP trials—especially when paired with HTP data such as NDVI, CT, and HSI, along with genomic information, breeders can enrich for superior lines and eliminate low performers earlier in the cycle. This enhances selection intensity and operational efficiency, particularly in the F5–F7 generations where replication is limited and seed quantity may restrict extensive BP evaluation. The scalability and non-destructive nature of UAV-based HTP platforms enable rapid, dynamic data collection across large populations, allowing for cost-effective prioritization of candidate lines [37]. Although this study did not focus on identifying specific SNP markers, the demonstrated predictive value of genomic features highlights opportunities for follow-up analyses, such as SNP effect estimation or variable importance mapping, which may guide future marker-assisted selection efforts. Overall, while BP trials remain essential for final variety evaluation, SP trials, when augmented with HTP and machine learning models, offer a practical and scalable tool for accelerating genetic gain earlier in the breeding pipeline.

5. Conclusions

Through this research, the effectiveness of combining SP HTP and genetic data to predict GY in BP trials has been explored. Through the utilization of SNP markers and contemporary HTP data collection and integration techniques, namely NDVI and CT, and also HSI, accurate yield projections can be made earlier in the breeding cycle, thus presenting considerable opportunities to save time and expenses while enhancing selection intensity. Though BP size trials are the benchmark for yield prediction owing to their extensive environmental scope, the smaller plot trials, when integrated with HTP data and ML models, potentially could be a cost-efficient option for earlier selection. This method shows the potential to further optimize wheat breeding and expedite the creation of high-yield, hardy cultivars.

Although genomic and HTP data collection can involve additional investment, the potential savings from reducing the scale of downstream BP trials may offset these costs, particularly when early-stage decisions are improved. The intent of this framework is not to suggest that all breeding programs must adopt every data stream simultaneously,

but rather to demonstrate how different data sources can be leveraged, individually or in combination, to improve predictive accuracy depending on available resources. For instance, in programs with limited access to hyperspectral imaging, models based on genomic and NDVI data alone still offer meaningful predictive value. Ultimately, the approach outlined here supports a flexible, modular strategy that breeding programs can adapt to their needs, allowing them to make evidence-based decisions about how and when to invest in more advanced phenotyping or genotyping tools.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy15061315/s1, Table S1: Genotypes used in the study and their respective states of origin. This table provides detailed information about the genotypic panel evaluated, including advanced experimental lines sourced from diverse breeding programs across several southern U.S. states. These genotypes represent a broad range of genetic variation and environmental adaptation; Figure S1: Thermal conditions throughout the growing season in Citra, FL. Temperature graph in degrees Celsius displaying the highest (green), average (blue), and lowest (orange) temperatures recorded every day for two consecutive years. This data set covers the period from November to May for the years 2021–22 and 2022–23; Figure S2: Rainfall in centimeters (cm) throughout the 2021–2022 and 2022–2023 growing seasons from mid-November to May of each year. The 2022 rainfall data is shown in blue while 2023 is in orange; Figure S3: Phylogenetic tree depicting the genetic relationships among the wheat breeding lines based on genome-wide SNP data. The tree was generated in TASSEL using a neighbor-joining clustering method based on pairwise genetic distance calculated from identity-by-state values.

Author Contributions: J.M. and M.A.B. conceptualized the research; Data curation was performed by J.M.; while formal analysis was carried out by J.M. and D.J.; funding acquisition was led by M.A.B.; investigation efforts involved J.M., N.K., S.K., J.P.A. and S.A.; methodology was developed by J.M. and M.A.B.; with project administration managed by M.A.B.; resources were provided by M.A.B., D.J., Y.A. and G.B.-G.; and software support was offered by D.J.; supervision was conducted by M.A.B., D.J. and Y.A.; visualization and original draft preparation were handled by J.M.; All authors, including J.M., M.A.B., D.J., Y.A., N.K., S.K., J.P.A., S.A. and G.B.-G. contributed to the review and editing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by UF/IFAS and the World Food Crops Breeding Program at UF, Project # 00123842.

Data Availability Statement: Data are available at: DOI: 10.5061/dryad.p5hqbzkzr; DOI: 10.5061/dryad.wwpzgmsvk.

Acknowledgments: We would like to express their sincere gratitude to the University of Florida's Plant Science Research and Education Unit (PSREU) for providing the facilities and resources essential to conducting this research. Special thanks are given to the SunGrains™ cooperative breeding program and its affiliated institutions, including the University of Arkansas, Clemson University, the University of Georgia, Louisiana State University, North Carolina State University, and Texas A&M University, for their contributions to the wheat genotypes used in this research. We also acknowledge the University of Florida's Institute of Food and Agricultural Sciences (UF/IFAS) for their funding and support.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BLUE	Best linear unbiased estimator
BP	Big plot

CT	Canopy temperature
CV	Cross validation
DAPCs	Discriminant analysis of principal components
DTH	Days to heading
G	Genomic data
GBR	Gradient boosting regression
GEBV	Genomic estimated breeding value
G×E	Genotype by environment interaction
GY	Grain yield
H	Hyperspectral data
HSI	Hyperspectral imaging
HTP	High throughput phenotyping
KNN	K-nearest neighbors
MAF	Minor allele frequency
MLR	Multiple linear regression
NDVI	Normalized difference vegetation index
RF	Random forest
RFR	Random forest regression
SNP	Single nucleotide polymorphism
SP	Small plot
SVMR	Support vector machine regression
UAV	Uncrewed aerial vehicle

References

1. Giraldo, P.; Benavente, E.; Manzano-Agugliaro, F.; Gimenez, E. Worldwide research trends on wheat and barley: A bibliometric comparative analysis. *Agronomy* **2019**, *9*, 352. [CrossRef]
2. Dohlman, E.; Hansen, J.; Boussios, D. Factors limiting the rate of dry matter in the grain of wheat grown at high temperature. *Aust. J. Plant Physiol.* **2022**, *7*, 121–140.
3. Erenstein, O.; Jaleta, M.; Mottaleb, K.A.; Sonder, K.; Donovan, J.; Braun, H.J. Global trends in wheat production, consumption and trade. In *Wheat Improvement: Food Security in a Changing Climate*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 47–66.
4. Zhu, Z.; Cao, Q.; Han, D.; Wu, J.; Wu, L.; Tong, J.; Xu, X.; Yan, J.; Zhang, Y.; Xu, K.; et al. Molecular characterization and validation of adult-plant stripe rust resistance gene Yr86 in Chinese wheat cultivar Zhongmai 895. *Theor. Appl. Genet.* **2023**, *136*, 142. [CrossRef]
5. Mittal, S. Wheat and barley production trends and research priorities: A global perspective. In *New Horizons in Wheat and Barley Research: Global Trends, Breeding and Quality Enhancement*; Springer: Singapore, 2022; pp. 3–18.
6. Sun, H.; Ma, J.; Wang, L. Changes in per capita wheat production in China in the context of climate change and population growth. *Food Secur.* **2023**, *15*, 597–612. [CrossRef]
7. Langridge, P.; Reynolds, M. Breeding for drought and heat tolerance in wheat. *Theor. Appl. Genet.* **2021**, *134*, 1753–1769. [CrossRef]
8. Fischer, R.A.; Rebetzke, G.J. Indirect selection for potential yield in early-generation, spaced plantings of wheat and other small-grain cereals: A review. *Crop. Pasture Sci.* **2018**, *69*, 439–459. [CrossRef]
9. Shi, Y.; Thomasson, J.A.; Murray, S.C.; Pugh, N.A.; Rooney, W.L.; Shafian, S.; Rajan, N.; Rouze, G.; Morgan, C.L.S.; Neely, H.L.; et al. Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS ONE* **2016**, *11*, e0159781. [CrossRef]
10. Sun, J.; Poland, J.A.; Mondal, S.; Crossa, J.; Juliana, P.; Singh, R.P.; Rutkoski, J.E.; Jannink, J.-L.; Crespo-Herrera, L.; Velu, G.; et al. High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. *Theor. Appl. Genet.* **2019**, *132*, 1705–1720. [CrossRef]
11. Mishra, S. Emerging technologies—Principles and applications in precision agriculture. In *Data Science in Agriculture and Natural Resource Management*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 31–53.
12. Li, J.; Yang, J.; Li, Y.; Ma, L. Current strategies and advances in wheat biology. *Crop. J.* **2020**, *8*, 879–891. [CrossRef]
13. Tattaris, M.; Reynolds, M.P.; Chapman, S.C. A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Front. Plant Sci.* **2016**, *7*, 1131. [CrossRef]
14. McBreen, J.; Babar, M.A.; Jarquin, J.D.; Lyerly, J.; Mergoum, M.; Murphy, J.P.; Boyles, R.E.; Harrison, S.; Ibrahim, A.M.; Shakiba, E.; et al. Improving grain yield in wheat lines adapted to the southeastern United States through multivariate and multi-environment genomic prediction models incorporating spectral and thermal information. *Plant Genome* 2024, *in press*.

15. Crain, J.; Mondal, S.; Rutkoski, J.; Singh, R.P.; Poland, J. Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome* **2018**, *11*, 170043. [CrossRef] [PubMed]
16. Montesinos-López, O.A.; Montesinos-López, A.; Crossa, J.; de los Campos, G.; Alvarado, G.; Suchismita, M.; Rutkoski, J.; González-Pérez, L.; Burgueño, J. Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* **2017**, *13*, 4. [CrossRef] [PubMed]
17. Zhang, Z.; Qu, Y.; Liu, H.; Wang, X.; Li, Y.; Chen, J. Integrating high-throughput phenotyping and genome-wide association studies for enhanced drought resistance and yield prediction in wheat. *New Phytol.* **2024**, *233*, 1234–1245. [CrossRef] [PubMed]
18. Galán, R.J.; Bernal-Vasquez, A.-M.; Jebsen, C.; Piepho, H.-P.; Thorwarth, P.; Steffan, P.; Gordillo, A.; Miedaner, T. Hyperspectral reflectance data and agronomic traits can predict biomass yield in winter rye hybrids. *BioEnergy Res.* **2020**, *13*, 168–182. [CrossRef]
19. Krause, M.R.; González-Pérez, L.; Crossa, J.; Pérez-Rodríguez, P.; Montesinos-López, O.; Singh, R.P.; Dreisigacker, S.; Poland, J.; Rutkoski, J.; Sorrells, M.; et al. Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3* **2019**, *9*, 1231–1247. [CrossRef]
20. McBreen, J.; Babar, A.; Jarquin, D.; Ampatzidis, Y.; Khan, N.; Kunwar, S.; Acharya, J.P.; Adewale, S.; Brown-Guedira, G. Enhancing genomic-based forward prediction accuracy in wheat by integrating UAV-derived hyperspectral and environmental data with machine learning under heat-stressed environments. *Plant Genome* 2024, *in press*.
21. Sirsat, M.S.; Oblessuc, P.R.; Ramiro, R.S. Genomic prediction of wheat grain yield using machine learning. *Agriculture* **2022**, *12*, 1406. [CrossRef]
22. Zadoks, J.C.; Chang, T.T.; Konzak, C.F. A decimal code for the growth stages of cereals. *Weed Res.* **1974**, *14*, 415–421. [CrossRef]
23. Lozada, D.N.; Godoy, J.V.; Ward, B.P.; Carter, A.H. Genomic prediction and indirect selection for grain yield in US Pacific Northwest winter wheat using spectral reflectance indices from high-throughput phenotyping. *Int. J. Mol. Sci.* **2020**, *21*, 165. [CrossRef]
24. International Wheat Genome Sequencing Consortium (IWGSC); Mayer, K.F.; Rogers, J.; Doležel, J.; Pozniak, C.; Eversole, K.; Feuillet, C.; Gill, B.; Friebe, B.; Lukaszewski, A.J.; et al. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **2014**, *345*, 1251788.
25. Bates, D.; Maechler, M.; Bolker, B.; Walker, S. *Lme4: Linear Mixed-Effects Models Using Eigen and S4, R Package Version 1.1-7*; R Foundation for Statistical Computing: Vienna, Austria, 2016. Available online: https://CRAN.R-project.org/package=lme4 (accessed on 15 May 2021).
26. Jarquín, D.; Crossa, J.; Lacaze, X.; Du Cheyron, P.; Daucourt, J.; Lorgeou, J.; Piraux, F.; Guerreiro, L.; Pérez, P.; Calus, M.; et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* **2014**, *127*, 595–607. [CrossRef]
27. Costa, L.; McBreen, J.; Ampatzidis, Y.; Guo, J.; Gahrooei, M.R.; Babar, M.A. Using UAV-based hyperspectral imaging and functional regression to assist in predicting grain yield and related traits in wheat under heat-related stress environments. *Precis. Agric.* **2021**, *23*, 622–642. [CrossRef]
28. Hu, Y.; Knapp, S.; Schmidhalter, U. Advancing high-throughput phenotyping of wheat in early selection cycles. *Remote. Sens.* **2020**, *12*, 574. [CrossRef]
29. Moghimi, A.; Yang, C.; Anderson, J.A. Aerial Hyperspectral Imagery and Deep Neural Networks for High-Throughput Yield Phenotyping in Wheat. *arXiv* **2019**, arXiv:1906.09666. [CrossRef]
30. Kaur, S.; Kakani, V.G.; Carver, B.; Jarquin, D.; Singh, A. Hyperspectral imaging combined with machine learning for high-throughput phenotyping in winter wheat. *Plant Phenomics J.* **2024**, *7*, e20111. [CrossRef]
31. Rutkoski, J.; Poland, J.; Mondal, S.; Autrique, E.; Pérez, L.G.; Crossa, J.; Reynolds, M.; Singh, R. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3* **2016**, *6*, 2799–2808. [CrossRef]
32. Montesinos-López, O.A.; Herr, A.W.; Crossa, J.; Carter, A.H. Genomics combined with UAS data enhances prediction of grain yield in winter wheat. *Front. Genet.* **2023**, *14*, 1124218. [CrossRef] [PubMed]
33. Michel, S.; Löschenberger, F.; Ametz, C.; Pachler, B.; Sparry, E.; Bürstmayr, H. Combining grain yield, protein content and protein quality by multi-trait genomic selection in bread wheat. *Theor. Appl. Genet.* **2020**, *132*, 2767–2780. [CrossRef]
34. Koh, J.C.; Spangenberg, G.; Kant, S. Automated machine learning for high-throughput image-based plant phenotyping. *Remote Sens.* **2021**, *13*, 858. [CrossRef]
35. Winn, Z.J.; Amsberry, A.L.; Haley, S.D.; DeWitt, N.D.; Mason, R.E. Phenomic versus genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines. *Plant Phenomics J.* **2023**, *6*, e20084. [CrossRef]

36. Heffner, E.L.; Lorenz, A.J.; Jannink, J.L.; Sorrells, M.E. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* **2011**, *4*, 65–75. [CrossRef]
37. Gill, H.S.; Poland, J.A.; Fritz, A.K.; Behl, J.D. Multi-trait genomic selection improves the prediction accuracy of end-use quality traits in hard winter wheat. *Plant Genome* **2022**, *15*, e20331. [CrossRef] [PubMed]