



OPEN ACCESS

EDITED BY

Yi-Hong Wang,
University of Louisiana at Lafayette,
United States

REVIEWED BY

Bahman Panahi,
University of Tabriz, Iran
Aras Türkoğlu,
Necmettin Erbakan University, Türkiye

*CORRESPONDENCE

Zoran Nikoloski

✉ nikoloski@mpimp-golm.mpg.de

Klára Panzarová

✉ panzarova@psi.cz

[†]These authors have contributed
equally to this work

RECEIVED 15 August 2025

ACCEPTED 22 September 2025

PUBLISHED 14 October 2025

CITATION

Tietze H, Abdelhakim L, Pleskačová B,
Kurtz-Sohn A, Fridman E, Nikoloski Z
and Panzarová K (2025) Prediction of harvest-
related traits in barley using high-throughput
phenotyping data and machine learning.
Front. Plant Sci. 16:1686506.
doi: 10.3389/fpls.2025.1686506

COPYRIGHT

© 2025 Tietze, Abdelhakim, Pleskačová, Kurtz-
Sohn, Fridman, Nikoloski and Panzarová. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Prediction of harvest-related traits in barley using high-throughput phenotyping data and machine learning

Hans Tietze^{1†}, Lamis Abdelhakim^{2†}, Barbora Pleskačová²,
Ayelet Kurtz-Sohn³, Eyal Fridman³, Zoran Nikoloski^{1,4*}
and Klára Panzarová^{2*}

¹Bioinformatics Department, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany, ²PSI (Photon Systems Instruments), spol. s r.o., Drásov, Czechia, ³Institute of Plant Sciences, Agricultural Research Organization (ARO), Bet Dagan, Israel, ⁴Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

Developing crop varieties that maintain productivity under drought is essential for future food security. Here, we investigated the potential of time-resolved high-throughput phenotyping to predict harvest-related traits and identify drought-stressed plants. Six barley lines (*Hordeum vulgare*) were grown in a greenhouse environment with well-watered and drought treatments, and dynamically phenotyped using RGB, thermal infrared, chlorophyll fluorescence, and hyperspectral imaging sensors. A temporal phenomic classification model accurately distinguished between drought-treated and control plants, achieving high accuracy (classification accuracy ≥ 0.97) even when relying solely on predictors from the early drought response phase. Canopy temperature depression at the early stage and RGB-derived plant size estimates at the late stage emerged as key classification features. A temporal phenomic prediction model of harvest-related traits achieved particularly high mean R^2 values for total biomass dry weight (0.97) and total spike weight (0.93), with RGB plant size estimators emerging as important predictors. Importantly, prediction accuracy for these traits remained high ($R^2 \geq 0.84$) even when restricted to early developmental phase data, including the stem elongation stage. Models trained on pooled drought and control data outperformed single-treatment models and maintained high predictive power across treatments. Together, these findings highlight the value of integrating high-throughput phenotyping with temporal modeling to enable earlier, more cost-effective selection of drought-resilient genotypes and demonstrate the broader potential of phenomics-driven strategies for accelerating crop improvement under stress-prone environments.

KEYWORDS

high throughput phenotyping, machine learning, phenomic prediction, plant breeding, barley and drought stress

Introduction

Climate change influences agricultural productivity and negatively affects crop yield, making the breeding of resilient crop varieties essential. The development of such stress-resilient varieties is challenging due to the interaction between genotype and environment that shapes complex traits, like grain yield. As a result, enhancing breeding programs for resilient crops requires accurate yield prediction across diverse environments (Cooper and Messina, 2023). Developing predictive models that integrate diverse data sets, e.g., genomics along with spatiotemporal phenomics and enviromics, can support this goal by enabling more accurate prediction of crop phenotypes (Xu et al., 2022). In addition, the implementation of advanced breeding techniques demands the development and deployment of high-throughput phenotyping (HTP) platforms in breeding programs. The resulting data along with the computational and machine learning approaches can improve future yield performance and help in developing resilient crop varieties that can withstand a variety of stresses, typical of field conditions (Varshney et al., 2021).

HTP is one of the techniques that has transformed and accelerated plant breeding by enabling large-scale, rapid screening of different phenotypic traits of interest, including automated data acquisition and trait analysis (Song et al., 2021; Yang et al., 2020). The use of multi-imaging sensors is essential for the non-invasive and precise assessment of plant growth dynamics as well as physiological responses. This approach provides a comprehensive view of plant development, enabling the monitoring and assessment of plant performance and stress responses (Cai et al., 2020; Humplik et al., 2015). Many studies have investigated the effect of abiotic stress, including drought, aiming at identifying the key phenotypic traits and physiological mechanisms that enhance stress tolerance (Al-Tamimi et al., 2022; Chen et al., 2014; Findurová et al., 2023). However, the complex nature of genotype-by-environment interactions remains a major challenge and demands further investigation. Moreover, a better understanding of drought adaptation requires recognizing that the impact of stress on physiological traits linked to grain yield can vary depending on stress intensity, genotype susceptibility, and developmental stage (Khadka et al., 2020). Along these lines, advances in high-throughput and precision phenotyping techniques have contributed to improving the strategies for mitigating the adverse effects of drought stress on plants and enhancing their resilience and productivity (Farooq et al., 2024).

One of the main challenges in harnessing the potential of high-throughput data lies in the management and analysis to identify traits of interest and reveal plant responses to stress (Langstroff et al., 2022; Leonelli et al., 2017). Data generated at multiple spatial and temporal scales requires robust analytical pipelines capable of handling such complex phenotypic datasets (Tardieu et al., 2017). Moreover, in phenotyping studies focusing on stress response across developmental stages, models are often modified to capture the dynamic changes of plant response to stress over time (Li et al., 2020). Recent pioneering advances have facilitated the prediction of the dynamics of multiple traits given genetic markers alone (Hobby

et al., 2025). Thus, while the development of analytical pipelines that explicitly capture the temporal dynamics of stress response in plants is highly demanded, our study advances the state of the art by integrating HTP with temporal modeling of harvest-related traits to enable predictions across developmental stages.

Machine learning techniques play a transformative role in phenotypic data analysis by linking large, complex datasets to traits of interest (Singh et al., 2016). Combining image-based phenotyping with machine learning approaches has enabled the extraction of new insights from curated, annotated, high-dimensional data sets across various crops and stress conditions (Singh et al., 2021). Machine learning encompasses a range of techniques, including feature extraction, pattern recognition, classification, and prediction. Some of these approaches facilitate the analysis of complex phenotypic data sets by considering multiple traits simultaneously, accounting for trait integration (Mbebi et al., 2025). As such, applying machine learning to phenomic data provides a powerful framework for uncovering patterns and extracting biologically meaningful insights (Gill et al., 2022). In this study, we focus on two widely used approaches: Random Forests, applied for both classification and regression, and least absolute shrinkage and selection operator (LASSO) regression, which provides a linear, regularized framework well suited to high-dimensional predictor sets.

Using the HTP platform equipped with multiple imaging sensors, we aimed to develop an advanced data analysis pipeline and apply it to perform a phenotypic data analysis of different barley (*Hordeum vulgare*) lines exposed to drought stress. We focused on barley as it is a model cereal crop (FAO, 2023; Newton et al., 2011), and we aimed to investigate the impact of drought as a predominant stress in future climate scenarios (IPCC, 2021). This was achieved by (i) using a classification model to identify distinct traits that differentiate drought-stressed from well-watered plants and (ii) using regression models to accurately predict harvest-related traits. The applied modeling approach enabled pinpointing the most predictive traits at specific time points. Moreover, early detection of such traits can support breeders in selecting stress-tolerant genotypes more efficiently, potentially accelerating the development of resilient crop varieties and improving resource use in breeding programs.

Materials and methods

Plant material and growth conditions

Six genetically homogenous barley lines were selected in this study, including one elite cultivar line (Barke) (L1) and five lines originating from the CMPP (Cytonuclear MultiParent Population) (L2-L5) and HEB-25 (Halle exotic barley) (L6) populations (HÜbner et al., 2009) (Supplementary Table S1).

After seeds were stratified at 4 °C in darkness, seedlings were transferred to light in the walk-in chamber (FytoScope FS-WI, Photon Systems Instruments (PSI), Drásov, Czech Republic) and were grown under a short-day regime, until the emergence of the

fifth leaf. One seedling was transplanted per 3-L pot filled with 1,850 g of Klasmann Substrate-2:sand (3:1). Plants were transferred to the greenhouse under a long day regime (16-h photoperiod), 22 ± 3 (mean $\pm$ standard deviation)/ 17 ± 2 °C for day/night temperature, and $51 \pm 8/62 \pm 4\%$ for day/night relative humidity.

Phenotyping protocol

The experiment was conducted in a greenhouse that is connected to the PlantScreenTM Modular phenotyping platform (PSI, Czech Republic), where pots were placed on transportation disks carried from the growth area toward the multi-imaging and irrigation units. Plant performance, including morphological and physiological responses, was assessed throughout the whole life cycle with an overall duration of plant cultivation of 97 days after transfer to light (DAT) and kept until reaching the full maturation stage (126 DAT). Over the course of 10 weeks, the daily phenotyping protocol was conducted to extract morpho-physiological and spectral-related traits in plants cultivated in semi-controlled greenhouse conditions under two watering regimes, control and progressive drought stress regime. Pots were weighed and watered daily by maintaining pots at target soil relative water content (SRWC) levels based on pot weight. Drought-stressed plants were maintained at 25% SRWC until the flowering stage, and then watering was further reduced to 20% SRWC (Supplementary Figure S1). We used nine biological replicates per treatment for most of the lines, and 20 replicates per treatment for the HEB line and elite line (Barke), which served as the reference line. The reduced watering regime was induced at the tillering stage (24 DAT) and remained reduced for the stressed group for the whole cultivation period. On a daily basis, plants were randomized in the cultivation greenhouse to avoid positional effects, environmental conditions were recorded with minute resolution, and daily watering and weighing of the plants were performed. Plants were phenotyped daily up to the maturity stage using multi-imaging sensors of the PlantScreenTM Modular phenotyping platform (PSI, Czech Republic), including RGB, thermal infrared (IR), chlorophyll fluorescence, and hyperspectral imaging, as described in Abdelhakim et al. (2024). Referring to chlorophyll fluorescence imaging, different measuring protocols were selected for capturing more insights into the photosynthetic performance, including a morning protocol and two different evening protocols (Supplementary Figure S1). During the day (light-adapted state), measuring protocols were optimized to measure the quantum yield of PSII (QY_{Lss}) under two light steady-state (Lss) intensities, including high light (HL, Lss1 at $1,200 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) and low light (LL, Lss2 at $130 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$). To estimate the plasticity index of QY under different light intensities, the ratio between QY_{Lss} measured under low (Lss2) to high (Lss1) light was calculated. Moreover, measurement on dark-adapted plants was conducted to assess the photosynthesis induction and relaxation kinetics during the night period at two different-light-intensity protocols, i.e., high

light (HL) at $1,200 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ and conditional light (CL) at $360 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ (Figure 1). At the end of the maturation stage, the total biomass of the plants was manually harvested, including analysis of the total tiller number, spike number, and other spike-related traits (Supplementary Table S2).

Data processing pipeline

The gathered dataset consists of dynamic phenotypical data from 70 time points captured for the six barley lines that were grown under two conditions, including 9–20 replicates per treatment per line. Overall, 145 image-based and post-harvest traits were extracted (Supplementary Table S2) and subjected to further data analysis. Of these, 52 traits from chlorophyll fluorescence and thermal IR imaging were excluded from downstream analysis, as they represented raw measurements used solely in the calculation of more biologically meaningful derived indices. The full data analysis pipeline was conducted using R studio (version 4.3.2).

Due to differing assumptions about input data across methods, preprocessing followed multiple branching paths. For analyses that separated drought and control treatments (harvest prediction per treatment), data were partitioned before outlier detection and transformation. In contrast, pooled-treatment analyses (i.e., variance decomposition, treatment classification, pooled-treatment harvest prediction) preserved treatment-induced variance by avoiding such partitioning. For temporal traits, each measurement time point was treated as a separate data group.

To maximize sample size and model robustness for the genotype-agnostic methods, including temporal phenomic classification (TPC) and prediction (TPP), three additional genotypic lines (L7-L9) were used in part of the analysis pipeline with the six genetically homogeneous lines (L1-L6) (Supplementary Table S1). However, as a result of heterogeneity, those three lines were excluded from analyses that assumed genetic homogeneity (i.e., clustering of samples, drought tolerance ranking, and trait variance decomposition).

Handling of missing values

Data preprocessing began with the imputation of 10 missing values across 267 plant samples in harvest traits, which were later used as response traits in TPP. Missing values were found in spike weight (nine samples) and total biomass dry weight (one sample). These were evenly distributed among samples, except for one sample (L8_D_15, Drought) with missing values in two traits. Missing values were imputed using MissForest imputation (Stekhoven and Bühlmann, 2011), implemented using the R package missForest (version 1.4), leveraging the remaining harvest traits. The out-of-bag (OOB) error for these imputations is provided in Supplementary Table S3.

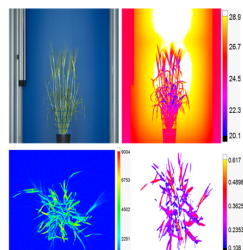


A. Experimental design

- 6 Barley lines
- 2 Treatments:
 - Control
 - Drought stress

B. Phenotyping protocol per week

- 3x ¹FC, IR, RGB1, RGB2
- 2x ²FC
- 1x ³FC, HS
- Daily weighing & watering



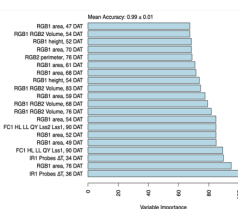
C. Image segmentation

Using Plant data analyzer software

- mask application
- subtraction of background
- pixel-by-pixel integration

D. Trait extraction

- From tillering to maturity stage
- Data set of 145 traits
- Across 70-time points



E. Data analysis

Machine learning approach

- Tolerance ranking
- Treatment prediction
- Harvest trait prediction

FIGURE 1

Overview of the experimental design and measurements performed. (A) Summary of the experimental design, including six barley lines. Two different water regime treatments were applied, control and drought stress at 60% and 25%-20% soil relative water content, respectively. Phenotyping was conducted from the tillering stage to the maturity stage, followed by final harvest. (B) Automated image-based phenotyping using the PlantScreen™ Modular phenotyping platform at PSI Research Center, where plants are moved from the greenhouse growing area toward imaging units. The phenotyping protocol was conducted daily with different protocols. In chlorophyll fluorescence imaging using FluorCam (FC), ¹FC morning measurement, ²FC night measurement, and ³FC chlorophyll content were conducted, as well as thermal infrared imaging (IR), RGB including two angles from the RGB1 side view and one angle RGB2 top view, and hyperspectral imaging (HS) including SWIR and VNIR imaging. (C) Automated image segmentation process. (D) Extracting traits, including measured and calculated parameters, among the developmental stages. (E) Data analysis using a machine learning approach to assess tolerance of plants under drought stress, discriminating between the two water regimes and finding the most predictive traits of the final yield.

Outlier handling

Outliers were identified as data points exceeding three times the interquartile range (IQR) of a given data group. These values were removed and re-imputed using missing forest imputation. The OOB errors for these imputations were reported ([Supplementary Table S4](#)). At this stage, the processed data were exported for variance decomposition.

Data transformation

A Shapiro–Wilk test for normality was applied to every data group, and p-values were corrected using the Bonferroni method. Those groups whose distribution was deemed non-normal had a Box–Cox transformation and were tested for normality again. Cases of non-normality before and after correction were reported ([Supplementary Table S5](#)). Following Box–Cox transformation of some groups, a Z-score transformation was applied to all groups. At this point, the transformed data were used for treatment classification and harvest trait prediction.

Week-wise aggregation of predictors

The same preprocessing steps used with the non-aggregated data set were also employed with the aggregated data set, including group-wise Random Forest imputation using missForest, outlier detection based on the IQR with a threshold of 3, re-imputation of extreme values after their removal, and transformation of non-normal trait distributions using Box–Cox followed by z-score normalization. Branching preprocessing paths were also applied, where treatment-specific analyses were conducted on partitioned data, while pooled-treatment analyses preserved treatment-induced variance by processing all samples jointly. A key difference lies in the temporal structuring of the data, whereas the original pipeline treated each measurement daily time point (DAT) as a separate data group; this pipeline uses weekly phases (WP) for grouping and aggregation. This approach reduces temporal noise while maintaining biological resolution, particularly relevant for trait dynamics across stress and recovery phases. The mapping between the DAT and the corresponding WP was defined in [Supplementary Table S6](#). For each numeric variable, the average (mean), minimum, and maximum values were calculated. To reduce redundancy, if all three values were identical within a group, indicating no variation, the minimum and maximum columns were removed, leaving only the average as the sole predictor.

Clustering of samples

All temporal traits from weekly aggregated data were combined, and principal component analysis (PCA) was performed using two R packages, prcomp and PCAtools, with scaled and centered data to

explore the underlying structure. Unsupervised clustering using k-means was then applied to the scaled trait data. Trait means were computed for each genotype-treatment combination and scaled. The optimal number of clusters was determined using the silhouette method with the R package factoextra (fviz_nbclust function). Clustering results were visualized using the fviz_cluster function and projected onto PCA space. Clusters were annotated by genotype and treatment, with PCA coordinates overlaid with confidence hulls to enhance interpretability.

Drought tolerance ranking of lines

The barley lines were ranked based on the magnitude of drought-induced effects on phenotypic traits. A permutational multivariate analysis of variance (PERMANOVA) with 3,000 permutations was applied separately to temporal and harvest traits, comparing drought-treated and control plants within each genotype ([Anderson, 2017](#)). This analysis was conducted using the R package vegan (version 2.6-8). Generalized eta squared (η^2) was used to quantify the treatment effect size, providing a measure of how strongly drought influenced each genotype's temporal or harvest traits.

Temporal phenomic classification of treatment

Random Forest (RF) binary classification of plant treatment was performed using temporal traits as predictors. Models were trained using the R packages caret (version 6.0-94) and randomForest (version 4.7-1.1). Each trait at each time point was used as an independent predictor. Models were trained using a threefold five-repeat cross validation (CV) scheme. In each repeat, data points were split randomly into three folds. For each split, two folds were used for training the model, and the last was reserved for testing. This was repeated three times, leaving each fold out for testing once.

To optimize model performance, the number of predictors randomly selected at each decision tree split was treated as a tunable parameter. A range of candidate values was systematically evaluated, and model performance was assessed across the 15 different test sets (three folds $\times$ five repeats) to ensure robustness. The final selection was based on the average performance across all repetitions, with the best-performing setting chosen to balance accuracy and model complexity.

Temporal phenomic classification (TPC) of treatment was conducted using temporal traits from all time points ($n_{\text{predictors}} = 850$) and using predictors from each separate week of the experiment ($53 \leq n_{\text{predictors}} \leq 92$, depending on the week). Model accuracies were reported as classification accuracy (i.e., the proportion of correctly classified samples out of the total number of samples tested) and compared using one-way ANOVA and a pairwise t-test with Holm's method for p-value correction.

To assess the variable importance of predictors, a permutation-based approach was used. For each tree in the forest, the

classification accuracy was first recorded using the out-of-bag (OOB) data, which consists of observations left out during bootstrap sampling. Then, the values of a given predictor were randomly permuted in the OOB data, and classification accuracy was re-evaluated. The drop in accuracy due to this permutation, relative to the original OOB accuracy, was computed for each tree. This accuracy difference was averaged across all trees, normalized by the standard deviation of the differences, and then scaled so that the most important variable received an importance score of 100.

To test whether the validation procedure affected the results, we also performed a leave-one-line-out CV. In this scheme, the models were trained while leaving out all replicates of one genetic line and then tested on that unseen line (nine validations). This directly evaluated model performance on previously unseen genetic lines.

Temporal phenomic prediction of harvest traits

Temporal phenomic prediction (TPP) of harvest traits was performed using least absolute shrinkage and selection operator (LASSO) with R package *glmnet* (version 4.1-8) and RF with R package *randomForest* (version 4.7-1.1) regression models. Training was done using the R package *caret* (version 6.0-94). The internal CV schedule was largely equivalent to the one used in TPC, except that LASSO models were optimized for the regularization strength parameter, rather than the number of variables considered at each split. In addition, the threefold CV procedure was repeated 15 times instead of 5 (3 folds $\times$ 15 repeats = 45 validations). The increased number of repeats was chosen based on preliminary testing, which showed greater variability in performance estimates for TPP models compared with TPC. For parameter optimization, model performance was evaluated using root mean squared error (RMSE). Finally, after the internal CV procedure had determined the optimal parameter, a final model was trained on the full predictor data set.

Separate models were trained to predict each of the 13 harvest traits using data from the control treatment ($n_{\text{plants}} = 133$), drought treatment ($n_{\text{plants}} = 134$), or a pooled dataset containing both treatments ($n_{\text{plants}} = 267$). Each temporal trait at each time point was treated as an independent predictor.

To assess whether the choice of validation strategy influenced results, TPP was also repeated using a leave-one-line-out CV scheme described in the TPC section. This procedure directly evaluated the robustness of harvest trait prediction when applied to genetic lines not included during training.

Harvest trait prediction models were trained using all traits measured at all time points ($n_{\text{predictors}} = 850$) and using only measurements from the first half of the experiment ($n_{\text{predictors}} = 368$). Model performance was compared using R^2 values originating from the repeated internal CV using the final optimal parameter, with statistical significance assessed using either ANOVA followed by multiple pairwise t-tests or a Kruskal–Wallis test followed by

multiple pairwise Wilcoxon rank-sum tests, depending on the normality of R^2 distributions. To correct for multiple testing, Holm's method was applied to adjust p-values.

Final models were tested on their training data set as well as the other data groups after readjusting them to the fitting z-transformation.

Variance decomposition of traits

Mixed effects linear models were used to model the temporal and harvest traits. Temporal and harvest traits were modeled similarly, except that the model used for temporal traits (Equation 1) included a temporal term, which was not the case for the harvest trait model (Equation 2). In addition, for harvest traits, the genetic repeatability (GR) was estimated using the genetic and residual variance components (Equation 3).

$$Y_{ijk} = \mu + \beta_i + \delta_{i(k)} + \Omega_{i(j)} + \varepsilon_{ijk} \quad (1)$$

where

Y temporal trait value

μ overall mean trait value for all plants

β_i random effect of the i th measuring time point
 $\sim N(0, \sigma^2_{\beta_i})$

$\delta_{i(k)}$ random effect of the treatment within i th time point
 $\sim N(0, \sigma^2_{\delta_{i(k)}})$

$\Omega_{i(j)}$ the random effect of j th genetic line within i th time point
 $\sim N(0, \sigma^2_{\Omega_{i(j)}})$

ε_{ijk} pooled error of the individual at i th time point
 $\sim N(0, \sigma^2_{\varepsilon_{ijk}})$

$$Y_{jk} = \mu + \delta_k + \Omega_j + \varepsilon_{jk} \quad (2)$$

where

Y harvest trait value

μ overall mean trait value for all plants

δ_k random effect of the k th treatment
 $\sim N(0, \sigma^2_{\delta_{i(k)}})$

Ω_j the random effect of j th genetic line
 $\sim N(0, \sigma^2_{\Omega_{i(j)}})$

ε_{jk} pooled error of the individual at i th time point
 $\sim N(0, \sigma^2_{\varepsilon_{ijk}})$

$$GR = \frac{\sigma^2_{\Omega_j}}{\sigma^2_{\Omega_j} + \frac{\sigma^2_{\varepsilon_{jk}}}{n}} \quad (3)$$

where

GR The genetic repeatability of a harvest trait

$\sigma^2_{\Omega_j}$ The genotypic variance

$\sigma^2_{\varepsilon_{jk}}$ The residual variance

n The number of replications

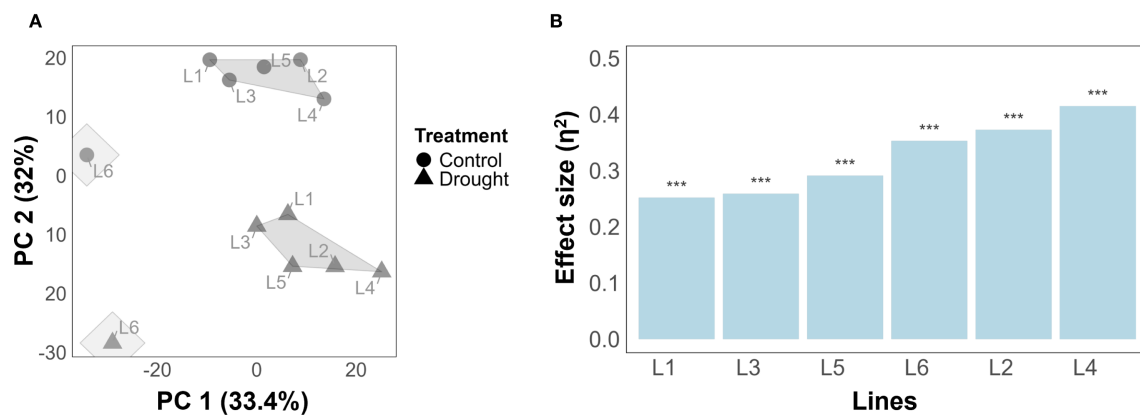


FIGURE 2

Treatment separation and ranking of the barley lines according to their susceptibility to drought stress. (A) k-means clustering using a two-dimensional PCA based on all traits across all time points for the six barley lines. Treatments are represented as control (circle shape) and drought (triangle shape). (B) Tolerance ranking of genotype to drought stress using PERMANOVA quantifying significance and effect size of treatment on temporal traits. The asterisks represent a significance level P -value < 0.001.

Results

Genotypic clustering and drought tolerance ranking of barley lines

To visualize the genotypic and treatment-specific grouping patterns, k-mean clustering (with $k = 2$) was applied on PCA-projected data. The results showed that separation of clusters can be observed based on PC1 (36.6%) and PC2 (31.3%) (Figure 2A). Notably, among the genetic lines, L6 (from HEB population) clustered separately from the other lines, highlighting potential differences in response to treatment conditions. Moreover, plants grown under control conditions were separated from drought-stressed plants, highlighting that treatment effects contribute to variance in the data.

Tolerance ranking, i.e., quantifying the effect of drought on temporal or harvest traits, found that drought has a highly significant effect ($p < 0.001$) on temporal traits in all lines (Figure 2B). By effect size, L1 appeared most tolerant in temporal traits ($\eta^2 = 0.228$), closely followed by L3 ($\eta^2 = 0.248$) and L5 ($\eta^2 = 0.274$). The same three lines appeared most tolerant when harvest traits were analyzed instead of temporal traits (Supplementary Figure S2), with lines L2, L4, and L6 exhibiting the least amount of tolerance with respect to both harvest and temporal traits.

Temporal phenomic classification of treatment

Temporal phenomic classification (TPC) of plant treatment (drought vs. control) was performed using all temporal predictors from daily time points across the full duration of the experiment (non-aggregated predictors) where model mean classification accuracy was 0.99 (Figure 3A). Using leave-one-line-out validation instead of the repeated threefold CV resulted in a

similarly high mean accuracy of 0.983. In another approach, predictors were aggregated by week using summary statistics (mean, minimum, and maximum), as described in the methods section, resulting in a slightly lower but maintained high mean accuracy of 0.98 compared with daily time points (Figure 3B). Variable importance analysis revealed that canopy temperature depression (ΔT) at the early stage (3 weeks of drought stress), along with, as expected, RGB-based plant size estimates (area from the RGB side view and plant volume) at the late stage (longer duration of drought stress) were the most influential predictors of treatment classification, regardless of daily or weekly aggregated data (Figure 3).

In addition, TPC using daily (non-aggregated) predictors was also performed by training separate models using only the data from individual weeks. All week-specific models, except the one based solely on week 0 (during which the drought treatment was initiated), achieved high accuracy in distinguishing treatments ($0.973 \leq \text{mean accuracy} \leq 0.99$; Supplementary Figure S3). The model trained exclusively on week 0 data had lower performance (mean accuracy = 0.695), likely due to the limited physiological response at this early stage. Notably, classification models using only the second and third weeks of measurements (corresponding to week 1 and 2 after inducing drought stress) relied almost entirely on ΔT estimates, along with traits from chlorophyll fluorescence and visible-near-infrared (VNIR) imaging (Supplementary Figure S4). This finding underscores the importance of these traits in early drought stress detection.

By comparing predictors derived from the daily and weekly aggregated data, we found that both contributed to identifying critical time points in the stress response, where common traits at specific time points were identified. Notably, from the chlorophyll fluorescence (morning protocol), the ratio of quantum yield under low to high light (FC1 HL LL QY Lss2/Lss1) was selected at day 90 and week 9, reflecting the severe impact of stress responses at the late stage (Figure 3).

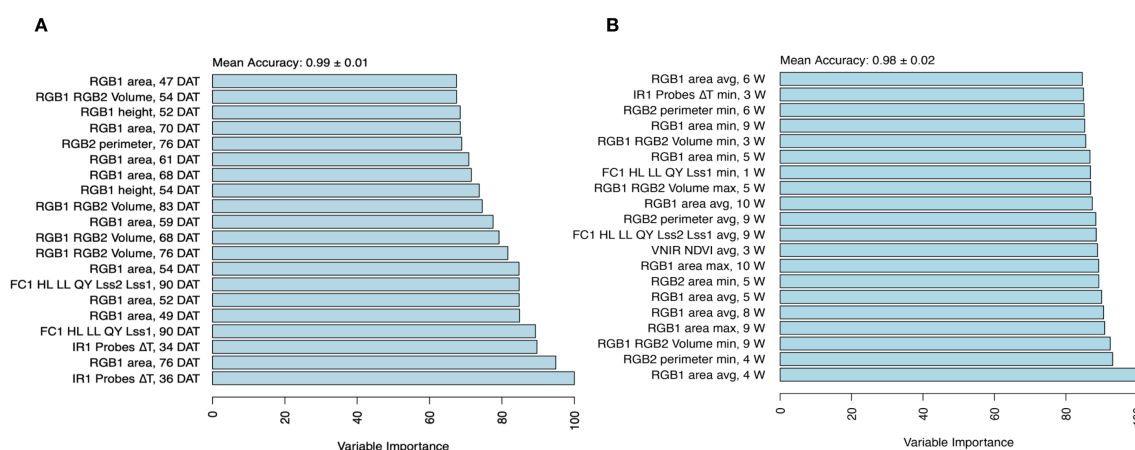


FIGURE 3

Variable importance for classification of treatment. The importance of the temporal traits was assessed using a random forest model with a mean accuracy metric for (A) non-aggregated data set and (B) aggregated data set composed of minimum, maximum, and average per week, and top 20 traits differentiating between treatments. Variable importance was determined using a permutation-based method, where the model's prediction accuracy was compared before and after shuffling each variable. Higher importance values indicate a greater decrease in accuracy when a variable is permuted, signifying its stronger contribution to classification performance.

Temporal phenomic prediction of harvest traits

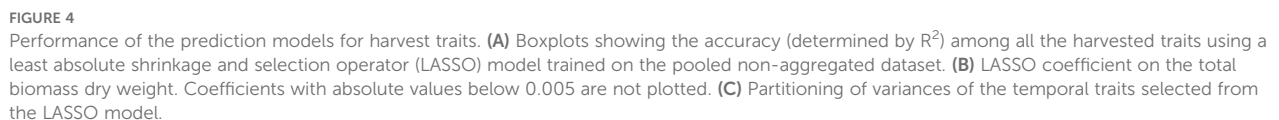
Temporal phenomic prediction (TPP) was also performed for all harvest traits using RF regression and LASSO models, with training data from the control treatment, drought treatment, or the pooled data set containing both treatments. Among the predicted traits, total biomass dry weight, spike number, total spike weight, and five spikes weight were generally the most predictable based on R^2 values (Figure 4A). Therefore, our subsequent comparisons of prediction accuracy focus on these four traits. Among these traits, total biomass exhibited substantially higher predictability compared with the other traits. Using leave-one-line-out validation instead of the repeated threefold CV resulted in similar mean accuracy values (pooled treatment LASSO results shown in Supplementary Figure S5), but larger variance, especially in poorly predictable traits. The high variance can likely be attributed to the smaller number of validation folds, the reduced sample size within each fold, and genuine differences in how well models generalize across individual genetic lines.

The relative performance of RF compared with LASSO depended on the training data set (i.e., control, drought, or the pooled treatment data) and the response trait, with only minor differences in overall accuracy between the two models (Supplementary Figure S6). Due to the small and unsystematic difference in performance between these model types, the rest of the findings reported will focus on the LASSO models, as they are facile to interpret without additional feature importance scoring. The choice of training data set (i.e., drought, control, or pooled) had a more pronounced effect on model performance than the choice of modeling approach. Models trained on the pooled dataset generally performed best, followed by those trained on drought data (Supplementary Figure S7). There is a slight overlap between useful predictors in pooled data TPP and TPC of treatment, suggesting that these models, by implicitly capturing treatment-related variance, leveraged the drought-induced variability

in both predictors and response traits to improve robustness (e.g., TPC, Figure 3A and biomass prediction, Figure 4B). This finding is further supported by the fact that many harvest traits, especially total biomass, spike weight, and spike number, showed a large variance contribution from treatment (Supplementary Figure S8). However, when using the pooled model to predict harvest traits using only drought or control data, the performance was marginally affected (Supplementary Table S7), indicating a reasonable degree of transferability of the model.

LASSO coefficients for pooled and drought treatment biomass prediction were generally associated with plant size estimates (e.g., plant area) derived from RGB imaging at late time points (Figure 4B). These predictors not only had the largest absolute coefficient values but also showed high variance contributions linked to the treatment (Figures 4B, C and Supplementary Figure S9). Aggregating predictors by week caused a slight drop in model performance during TPP, but accuracies were still similar to those found in the original daily data set (Supplementary Figure S10).

To evaluate the feasibility of early prediction of harvest traits and to identify informative early predictors, models were retrained using only measurements from the first half of the experiment (early phase including stem elongation) (Supplementary Figure S11). Compared with models trained on all time points, R^2 prediction accuracy was decreased and had higher variance in all models. However, despite this reduction in accuracy, predictions remained reasonably effective, particularly for biomass dry weight (mean $R^2 = 0.924$ for pooled-treatment LASSO model) and total spike weight (mean $R^2 = 0.837$ for pooled-treatment LASSO model) (Supplementary Figure S12). In these early-timepoint models, pooled-treatment models still resulted in the highest overall accuracy. RGB plant size estimates were again found to be important predictors, but their contribution was complemented by predictors from other sensors. Biomass prediction incorporated hyperspectral indices (MCARI1, LWVI2, NDVI) (Supplementary Figure S12A), whereas total spike weight relied more strongly on photosynthetic efficiency from chlorophyll fluorescence, supplemented



accuracy was consistently high (mean accuracy ≥ 0.97) when using temporal predictors from any week following drought initiation, with reduced performance only during the initial week when stress responses were not yet fully established.

Model accuracy and cross validation

An important outcome of the harvest trait prediction analysis was the clear superiority of models trained on the pooled dataset. This increase in accuracy was likely due to structured variability in temporal and harvest traits introduced by treatment, as can be seen in the substantial treatment-induced variance in predictors and

response traits, as well as in the overlap in important predictors from TPC and pooled treatment TPP. However, the increased sample size used to train pooled models should also be acknowledged as a contributing factor. In addition to achieving the highest accuracy across internal cross-validation folds, pooled treatment models also performed well when tested separately on drought and control treatment data (Supplementary Table S7). While this test does not constitute a true external cross-validation, it provides evidence that the improved performance of pooled models is not merely driven by treatment-level separation. The high accuracy observed when predicting within each treatment group suggests that the models capture meaningful, continuous variation among individual plants, rather than just categorical differences between drought and control conditions.

External validation, in which a model is tested on data entirely excluded from training, remains the most rigorous method for identifying overfitting and assessing model generalization. While external validation was not feasible in this study, model performance was evaluated using repeated internal cross-validation, where R^2 values were computed on data withheld from training. This approach reduces the risk of overfitting. Apart from this, other observations are indicative that overfitting was not the cause for the high observed accuracy.

Most notably, prediction accuracy varied substantially across response traits despite the use of identical predictor sets. Traits with limited or indirect association to vegetative growth dynamics, such as infertile tip length and spikelet density, consistently yielded low predictive performance ($R^2 < 0.3$). If the models were overfitted, one would expect uniformly high accuracy across traits, regardless of biological relevance. Instead, this pattern suggests that the high accuracy observed for traits, such as biomass and spike weight, likely reflects a reliable predictive signal rather than a model artefact. A similar argument holds for the treatment classification task, whereas most weekly models achieved near-perfect accuracy, classification using only the first week of measurements performed poorly. This was biologically expected, as the drought treatment had only just started during this week (WP 0) and had not yet induced measurable phenotypic changes. The presence of both high- and low-performing models, in line with biological expectations, provides indirect evidence that overfitting is not the primary driver of the observed model performance.

Moreover, both regression methods using LASSO and Random Forest have inherent mechanisms for mitigating overfitting. LASSO applies L_1 regularization, which penalizes model complexity by shrinking coefficients and effectively selecting a subset of informative features (Ying, 2019). Random Forests, through ensemble learning and the use of out-of-bag error estimation, reduce the risk of overfitting by averaging across multiple decorrelated decision trees (Ghojogh and Crowley, 2023). Nonetheless, while these approaches help control overfitting, a more reliable way to assess generalizability is through external cross-validation using independent data not seen during model training or optimization (Ghojogh and Crowley, 2023). Due to the limited sample size and experimental design, such validation was not feasible in this study, but should be prioritized in future work.

Integrative insights into model interpretability, classification, and prediction

Variable importance was assessed using model coefficients in LASSO and a permutation-based approach in Random Forests. These measures provide insight into predictive relevance, but reflect conditional importance, which may fail to identify all relevant predictors when multicollinearity is present, as only one among a set of correlated variables may be selected or appear important (Grömping, 2009). This is likely the reason for the slight inconsistencies observed in variable importance rankings between models using daily (non-aggregated) and weekly (aggregated) predictors.

Temporal phenomic classification of treatment

For treatment classification, canopy temperature depression (ΔT) at early time points and RGB plant size estimates at late time points emerged as important predictors (Figure 3). Notably, treatment could be classified very accurately early in the experiment mainly by relying on ΔT measurements, highlighting the importance of this trait for early drought detection. Plant canopy temperature has an impact on plant growth by non-linearly regulating photosynthesis, respiration, and transpiration rates. By increasing water deficit, the efficiency of cooling the leaf surface through transpiration diminishes, leading to an increase in leaf temperature (Biju et al., 2018; Way and Yamori, 2014). Therefore, changes in canopy temperature provide a valuable proxy of stomatal regulation and an indicator of different stress responses (Wen et al., 2023). Stomatal closure serves as an initial response to drought stress to prevent excessive water loss, leading to alterations in physiological response and metabolic pathways (Farooq et al., 2024). In our study, drought-stressed plants showed higher ΔT due to increased canopy temperature. However, as drought duration increases over time, ΔT became a less significant classifier of treatment, likely due to physiological adaptations such as stomatal acclimation, osmotic regulation, and changes in transpiration dynamics, which can moderate canopy temperature despite continued water limitation (Lawson and Blatt, 2014). In addition, plants progressively reduce their biomass due to water limitations (Pham et al., 2019). This structural change, including leaf area reduction, alters canopy-atmosphere interactions and subsequently affects canopy temperature regulation (Vico et al., 2023).

Temporal phenomic prediction of harvest traits

For the prediction of harvest traits, RGB-based plant size estimates from late time points emerged as dominant predictors, particularly in models for biomass (Figure 4). This is likely due to the strong biological alignment between total biomass and plant size, as both reflect cumulative vegetative growth. This aligns with previous studies showing the reduction of biomass accumulation under drought stress (Cai et al., 2020; Neumann et al., 2015). Moreover, the high correlation between the biomass and yield, where biomass reduction ultimately affects yield-related traits, as a result of low assimilates for grain production, reflects a source-to-sink limitation (Findurová et al., 2023; Rosati and Benincasa, 2023). When prediction models for biomass and total spike weight were trained using only pooled-treatment predictors from the first half of

the experiment, RGB plant size traits remained influential, but they were complemented by features derived from all of the other sensors. Despite a significant reduction in predictive accuracy compared with models trained on the full dataset, these early-timepoint models still achieved strong performance, with mean R^2 values of 0.92 for biomass and 0.84 for spike weight, respectively (Supplementary Figures S11, S12). The relatively high accuracy of these early TPP models is particularly promising in a breeding context, as it demonstrates that complex cumulative traits can be reliably predicted well before they are phenotypically expressed. Harnessing this predictive capacity would allow breeders to select plants early in the breeding pipeline, significantly reducing cost.

While RGB showed importance and strong associations with harvest-related traits, integrating multiple sensors enriches the predictive power of the models and strengthens the biological interpretability. Previous studies highlighted the importance of multiple imaging sensors besides RGB for deeper insights into the mechanisms of stress resilience (Shi et al., 2025; Zhang et al., 2024). Notably, the fraction of open Photosystem II (PSII) centers at light steady state (qL_Lss) showed importance in LASSO coefficient when prediction models were trained on pooled and drought data during the early drought stage (stem elongation stage) (Figure 4, Supplementary Figure S9). The variation in the opening of the PSII centers is probably reflecting alterations in photosynthetic efficiency and electron transport pathways, which shows major mitigation mechanisms to alleviate the negative effects of moderate drought stress (Qiao et al., 2024; Shin et al., 2021). By increasing the stress intensity and duration, the quantum yield of PSII (QY Lss2/Lss1) was observed at the late stage as an important predictor, indicating impairment of PSII function, where plants were unable to efficiently transfer energy from high to low light (Zhou et al., 2019).

Implications for breeding and timely selection

One of the primary applications of TPP is enabling the early selection of plants in breeding. In this study, it was demonstrated that this task is feasible with high accuracy for a variety of traits. Even predictors from the early phase of the phenotyping period (including the stem elongation stage) in the models achieved high accuracy for traits such as total biomass and total spike weight, potentially allowing selection after only a few weeks of vegetative growth. Notably, we observed that complementary traits showed importance from the hyperspectral data contributing strongly to early-stage predictions, emphasizing the importance of hyperspectral imaging. Similarly, recent work has shown the importance of hyperspectral reflectance for predicting a wide range of physiological traits, applying advanced machine learning approaches, including different prediction scenarios (Xu et al., 2025). However, it should also be noted that models trained on predictors from the full phenotyping period remain valuable to breeders, as phenotyping concluded 31 days before harvest. Nevertheless, daily phenotyping and harvesting are both costly and time-consuming processes, which could be partially avoided by implementing early phenotyping to identify tolerant plants prior to the later stages of evaluation (Adak et al., 2023).

An increase in prediction accuracy was observed when data from both drought and control treatments were combined in the training

set. This illustrates a key advantage of phenomic prediction over genomic prediction. While genomic prediction captures only the static genetic contribution to trait variation, phenomic prediction leverages genetic, environmental, and genotype-by-environment interaction effects, as phenotypic traits, both predictors and responses, reflect the integrated output of all these factors (Adak et al., 2023). Variance decomposition of both predictor and harvest traits in this study further illustrates that trait variation arises from a combination of genetic background, environmental conditions, and temporal dynamics. In contrast, genomic prediction models are inherently limited to the genetic component of variance and cannot account for time-dependent or environmentally induced effects. The inclusion of a temporal dimension in phenomic prediction further enhanced model performance by capturing dynamic shifts in environmental conditions and their interactions with genotype over time. This ability to model temporal trajectories of plant development and stress response adds substantial predictive power and highlights the unique potential of time-resolved phenomic data in breeding applications.

The implications for breeding are complicated by the fact that this experiment was performed under greenhouse conditions. While field trials are essential for capturing the full environmental variance affecting crop performance, phenotyping data from controlled environments remain highly valuable. They enable preliminary screening, identification of heritable differences between genotypes, and the generation of high-quality training data for predictive crop modeling and genetic analyses (Debbagh et al., 2025; Rahaman et al., 2015; Rayaprolu et al., 2025). Controlled-environment studies also allow detailed investigation of stress responses and developmental dynamics, as well as the measurement of key traits that are difficult to capture reliably in the field due to environmental heterogeneity and technical constraints (Langstroff et al., 2022). Importantly, recent work has shown that controlled platforms are not isolated from real-world relevance; they can simulate weather conditions and produce growth dynamics comparable with those observed in the field (Heuermann et al., 2023). Thus, controlled environments provide a powerful complement to field phenotyping for trait dissection, breeding, and modeling under current and future climate scenarios.

Future perspectives

Viewed in a breeding context, while this study focused on yield-related traits such as total spike weight, spike number, and biomass provide meaningful proxies, they do not fully capture grain production in the field. Future work could incorporate direct grain yield measurements, which are typically the primary selection targets in breeding programs and ensure greater relevance. Nevertheless, the study offers valuable insights into the predictive capacity of temporal phenomic data, and demonstrates how yield-associated traits can be modeled early in the growth period. The approaches presented here are readily transferable to datasets including grain yield, and thus remain relevant for informing breeding strategies.

The present study selected a relatively low number of genotypes examined (six genetically homogeneous lines and three genetically

heterogeneous ones) to demonstrate methodological feasibility as expanding genetic diversity was not the scope of the study, besides the limited capacity of the phenotyping platform in screening the lines. This restricted diversity likely reduces the generalizability of the models and does not fully reflect the genetic variation encountered in breeding programs. Scaling up to a broader and more diverse genetic panel would increase the robustness of predictions and better mimic the setting of applied breeding.

As previously discussed, models were evaluated using a repeated threefold internal cross-validation procedure, which provides the most reliable estimate of predictive performance, given the available sample size. However, implementing external cross-validation for performance assessment in temporal phenomic prediction in future studies would offer a more robust measure of model generalizability and could be used to assess performance across different environmental conditions. Cross-environment validation has previously been applied in TPP studies to demonstrate its advantage over genomic prediction, which often does not transfer to new environments without a significant decrease in accuracy (Adak et al., 2023; Jarquin et al., 2021). Incorporating genomic prediction alongside the phenomic methods used in this study would also enable a direct comparison of their relative predictive power. However, this strategy would require a substantial increase in the phenotyped population and application of recently developed data integrative approaches (Hobby et al., 2025).

For TPC and TPP, prediction accuracy in the field would likely be reduced due to environmental variability, which introduces additional noise affecting both predictor and response traits. The key challenge in field studies is the technical difficulty of collecting the same high-frequency, multi-sensor trait data as in the greenhouse. Nevertheless, our results demonstrate that both TPC and TPP can be successfully applied under controlled conditions, providing a proof of concept. The next step is to expand these approaches and test this pipeline in field experiments, which would capture more realistic environmental variability and allow the inclusion of more genotypes and larger sample sizes. This would enable more rigorous validation schemes, such as external cross-validation. With larger genotype panels, prediction could also be framed at the genotype level, more closely reflecting how predictions are applied in breeding programs.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#). The data and analysis pipeline code is available on <https://github.com/hatiez/barley-TPP-pipeline>.

Author contributions

HT: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. LA: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. BP: Data curation,

Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing. AK-S: Conceptualization, Investigation, Methodology, Resources, Visualization, Writing – review & editing. EF: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. ZN: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Visualization, Writing – review & editing. KP: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This study in the framework of the CAPITALISE project was funded by the European Union's Horizon 2020 research and innovation program (grant no. 862201).

Acknowledgments

We acknowledge David Hobby for his feedback during the development of the analysis pipeline. We thank Pavla Homolová for helping prepare plant material during the experiment, Ivan Kashkan, Sajid Ullah for segmentation analysis and Jaromír Pytela for technical support at PSI Research Center, Czech Republic.

Conflict of interest

LA and BP were employed by Photon Systems Instruments.

The remaining 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1686506/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Overview of the experimental timeline and phenotyping protocol. Phenotyping was conducted from the tillering stage to the maturity stage, followed by final harvest for harvest-related traits assessment. The drought stress was induced at the tillering stage, where the pot weight was maintained at 25% soil relative water content (SRWC), and during the flowering stage, the intensity of drought stress increased at 20% SRWC. The phenotyping protocol was conducted daily with different protocols using the PlantScreen™ Modular phenotyping platform at PSI Research Center. In chlorophyll fluorescence imaging using FluorCam (FC), ¹FC morning measurement for the quantum yield of PSII (QY_Lss), ²FC night measurement using conditional and high light levels, ³FC morning measurement for chlorophyll content were conducted, thermal infra-red imaging (IR), RGB including two angles from RGB1 side view and one angle RGB2 top view, and hyperspectral imaging (HS) including SWIR and VNIR imaging. And daily weighing and watering (WW) to maintain the target weight.

SUPPLEMENTARY FIGURE 2

Ranking the lines according to their susceptibility to drought stress on harvest traits. PERMANOVA was used to quantify the significance and effect size of the treatment on harvest traits. The asterisks represent significance level P -value < 0.001 for *** and 0.01 for **.

SUPPLEMENTARY FIGURE 3

Temporal phenomic classification (TPC) of treatment (drought/control) using predictors subset by week. TPC was performed using random forest models. Predictors from each week were used separately along with a model trained on the full data set (all weeks).

SUPPLEMENTARY FIGURE 4

Variable importance of temporal phenomic classification (TPC) of treatment using predictors from two weeks. (A) Traits from week 1 and (B) week 2 after inducing drought stress. TPC was performed using random forest models.

SUPPLEMENTARY FIGURE 5

Boxplots showing the accuracy (determined by R^2) among all the harvested traits using the LASSO model trained using leave-one-line-out validation instead of the repeated 3-fold CV (9 validations instead of 45).

SUPPLEMENTARY FIGURE 6

Comparison between the accuracy of models. Boxplots showing the accuracy (determined by R^2) among selected harvested traits using LASSO and Random Forest models trained on a pooled non-aggregated dataset. The significance level was determined as *** for $P < 0.001$, **** for $P < 0.0001$ and ns for non-significant differences between the models.

SUPPLEMENTARY FIGURE 7

Comparison between the different treatments using the LASSO model. Boxplots showing the accuracy (determined by R^2) among selected harvested traits using the LASSO model trained on a non-aggregated dataset. The significance level was determined as **** for $P < 0.0001$ and ns for non-significant differences between the models.

SUPPLEMENTARY FIGURE 8

Partitioning of variances and repeatability of the harvest-related traits. Repeatability on the secondary y-axis is represented as error bars.

SUPPLEMENTARY FIGURE 9

Harvest trait prediction models on drought non-aggregated dataset. (A) Boxplots showing the accuracy (determined by R^2) among all the harvested

traits using the LASSO model trained on drought non-aggregated dataset. (B) LASSO coefficients on the total biomass dry weight where coefficients with absolute values below 0.005 are not plotted. (C) Partitioning of variances of the temporal traits selected from LASSO model.

SUPPLEMENTARY FIGURE 10

Harvest-traits prediction models on an aggregated dataset. (A) Boxplots showing the accuracy (determined by R^2) among all the harvested traits using the LASSO model trained on pooled aggregated dataset. (B) LASSO coefficient on the total biomass dry weight. (C) Partitioning of variances of the temporal traits selected from LASSO model.

SUPPLEMENTARY FIGURE 11

Comparison between the different early and all time points using the LASSO model. Boxplots showing the accuracy (determined by R^2) among selected harvested traits using the LASSO model trained on pooled non-aggregated dataset. Early time points were selected until 51 days after transplant (DAT). The significance level was determined as **** for $P < 0.0001$ and ns for non-significant differences between the models.

SUPPLEMENTARY FIGURE 12

LASSO coefficients of early time point TPP models for harvest traits. (A) Biomass dry weight and (B) total spike weight were selected. Only coefficients with absolute values above 0.015 are plotted.

SUPPLEMENTARY TABLE 1

Genetic background of the plant material for the selected lines.

SUPPLEMENTARY TABLE 2

List of temporal traits description from multiple imaging sensors and harvest-related traits.

SUPPLEMENTARY TABLE 3

Out of bag (OOB) error of imputed missing values in harvest traits. In all samples, missing values were only in the total spike weight trait, except in L8, where total biomass was also missing. OOB error is calculated as normalized root mean square error (NRMSE) based on predictions for observed values that were left out of the training bootstrap samples (i.e., out-of-bag) when fitting the random forest.

SUPPLEMENTARY TABLE 4

Outlier rate and re-imputation out of bag (OOB) error for each sensor type, including different protocols as described in Figure 1. Outliers were detected for each trait among all replicates within a genetic line within a treatment.

SUPPLEMENTARY TABLE 5

Rates of non-normality (NN) for each sensor type, including different protocols as described in Figure 1. Non-normality was tested for each trait within each data group, either separated by treatment or pooled across treatments. Groups identified as non-normal were corrected using a Box-Cox transformation, and NN rates were recalculated post-correction.

SUPPLEMENTARY TABLE 6

Mapping between the daily time point counting from days after transfer of seedlings to light (DAT) for the non-aggregated dataset and the corresponding weekly phases (WP) for the aggregated dataset. The developmental stage was marked when most of the replicates among all the genetic lines reached a certain stage. Early stage refers to the vegetative and early reproductive phases, during which plants were exposed to 3 weeks of drought stress. Late stage refers to the late reproductive phase, when the flag leaf was fully expanded, and plants were exposed to a longer duration of drought stress.

SUPPLEMENTARY TABLE 7

Performance of final LASSO models trained on the full predictor set using the optimal parameter (λ). Accuracies much higher than those found in the internal CV procedure during training suggest overfitting. "Training Group" refers to the data group used to train the model, whereas the R^2 values show the accuracy when the model is applied to the data group in question.

References

- Abdelhakim, L. O. A., Pleskačová, B., Rodriguez-Granados, N. Y., Sasidharan, R., Perez-Borroto, L. S., Sonnewald, S., et al. (2024). High throughput image-based phenotyping for determining morphological and physiological responses to single and combined stresses in potato. *JoVE* 208, e66255. doi: 10.3791/66255
- Adak, A., Murray, S. C., and Anderson, S. L. (2023). Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. *G3: Genes Genomes Genet.* 13 (1), jkac294. doi: 10.1093/g3journal/jkac294
- Al-Tamimi, N., Langan, P., Bernád, V., Walsh, J., Mangina, E., and Negrão, S. (2022). Capturing crop adaptation to abiotic stress using image-based technologies. *Open Biol.* 12, 210353. doi: 10.1098/rsob.210353
- Anderson, M. J. (2017). Permutational multivariate analysis of variance (PERMANOVA). In: *Wiley StatsRef: Statistics Reference Online*, N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri and J. L. Teugels, John Wiley & Sons, Ltd., Hoboken, NJ, 1–15. doi: 10.1002/9781118445112.stat07841
- Biju, S., Fuentes, S., and Gupta, D. (2018). The use of infrared thermal imaging as a non-destructive screening tool for identifying drought-tolerant lentil genotypes. *Plant Physiol. Biochem.* 127, 11–24. doi: 10.1016/j.plaphy.2018.03.005
- Cai, K., Chen, X., Han, Z., Wu, X., Zhang, S., Li, Q., et al. (2020). Screening of worldwide barley collection for drought tolerance: the assessment of various physiological measures as the selection criteria. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01159
- Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., et al. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis w open. *Plant Cell* 26, 4636–4655. doi: 10.1105/tpc.114.129601
- Cooper, M., and Messina, C. D. (2023). Breeding crops for drought-Affected environments and improved climate resilience. *Plant Cell* 35 (1), 162–186. doi: 10.1093/plcell/koac321
- Debbagh, M., Sun, S., and Lefsrud, M. (2025). Predictive modeling, pattern recognition, and spatiotemporal representations of plant growth in simulated and controlled environments: A comprehensive review. *Plant Phenomics* 7 (3), 1–18. doi: 10.1016/j.plaph.2025.100089
- FAO (2023). *Agricultural production statistics 2000–2022*. FAOSTAT Analytical Briefs, No. 79. Rome: FAO. doi: 10.4060/cc9205en
- Farooq, M., Wahid, A., Zahra, N., Hafeez, M. B., and Siddique, K. H. M. (2024). Recent advances in plant drought tolerance. *J. Plant Growth Regul.* 43 (10), 3337–3369. doi: 10.1007/s00344-024-11351-6
- Findurová, H., Veselá, B., Panzarová, K., Pytela, J., Trtílek, M., and Klem, K. (2023). Phenotyping drought tolerance and yield performance of barley using a combination of imaging methods. *Environ. Exp. Bot.* 209, 105314. doi: 10.1016/j.envexpbot.2023.105314
- Ghojogh, B., and Crowley, M. (2023). The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. *ArXiv* 1905, 12787. doi: 10.48550/arXiv.1905.12787
- Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., and Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics* 2, 156–183. doi: 10.1007/s43657-022-00048-z
- Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *Am. Statistician* 63, 308–319. doi: 10.1198/tast.2009.08199
- Heuermann, M. C., Knoch, D., Junker, A., and Altmann, T. (2023). Natural plant growth and development achieved in the IPK PhenoSphere by dynamic environment simulation. *Nat. Commun.* 14 (1), 5783. doi: 10.1038/s41467-023-41332-4
- Hobby, D., Tong, H., Heuermann, M., Mbebi, A. J., Laitinen, R. A. E., Dell'Acqua, M., et al. (2025). Predicting plant trait dynamics from genetic markers. *Nature*. 11, 1018–1027. doi: 10.1038/s41477-025-01986-y
- Hübner, S., Höfken, M., Oren, E., Haseneyer, G., Stein, N., Graner, A., et al. (2009). Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. *Mol. Ecol.* 18, 1523–1536. doi: 10.1111/j.1365-294X.2009.04106.x
- Humplik, J. F., Lázár, D., Husičková, A., and Spíchal, L. (2015). Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses - a review. *Plant Methods* 11, 1–10. doi: 10.1186/s13007-015-0072-8
- IPCC (2021). Climate Change 2021: The Physical Science Basis. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2391 pp. doi: 10.1017/9781009157896
- Jarquín, D., de Leon, N., Romay, C., Bohn, M., Buckler, E. S., Ciampitti, I., et al. (2021). Utility of climatic information via combining ability models to improve genomic prediction for yield within the genomes to fields maize project. *Front. Genet.* 11. doi: 10.3389/fgene.2020.592769
- Khadka, K., Earl, H. J., Raizada, M. N., and Navabi, A. (2020). A physiological trait-based approach for breeding drought tolerant wheat. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00715
- Langstroff, A., Heuermann, M. C., Stahl, A., and Junker, A. (2022). Opportunities and limits of controlled-environment plant phenotyping for climate response traits. *In Theor. Appl. Genet.* 135, 1–16. doi: 10.1007/s00122-021-03892-1
- Lawson, T., and Blatt, M. R. (2014). Stomatal size, speed, and responsiveness impact on photosynthesis and water use efficiency. *Plant Physiol.* 164, 1556–1570. doi: 10.1104/pp.114.237107
- Leonelli, S., Davey, R. P., Arnaud, E., Parry, G., and Bastow, R. (2017). Data management and best practice for plant science. *Nat. Plants* 3 (6), 1–4. doi: 10.1038/nplants.2017.86
- Li, Z., Guo, R., Li, M., Chen, Y., and Li, G. (2020). A review of computer vision technologies for plant phenotyping. *Comput. Electron. Agric.* 176, 1–21. doi: 10.1016/j.compag.2020.105672
- Mbebi, A. J., Mercado, F., Hobby, D., Tong, H., and Nikoloski, Z. (2025). Advances in multi-trait genomic prediction approaches: classification, comparative analysis, and perspectives. *Briefings in Bioinformatics*, 26 (3), 1–16. doi: 10.1093/bib/bba211
- Neumann, K., Klukas, C., Friedel, S., Rischbeck, P., Chen, D., Entzian, A., et al. (2015). Dissecting spatiotemporal biomass accumulation in barley under different water regimes using high-throughput image analysis. *Plant Cell Environ.* 38, 1980–1996. doi: 10.1111/pce.12516
- Newton, A. C., Flavell, A. J., George, T. S., Leat, P., Mullholland, B., Ramsay, L., et al. (2011). Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security. *Food Secur.* 3, 141–178. doi: 10.1007/s12571-011-0126-3
- Pham, A. T., Maurer, A., Pillen, K., Brien, C., Dowling, K., Berger, B., et al. (2019). Genome-wide association of barley plant growth under drought stress using a nested association mapping population. *BMC Plant Biol.* 19 (1), 134. doi: 10.1186/s12870-019-1723-0
- Qiao, M., Hong, C., Jiao, Y., Hou, S., and Gao, H. (2024). Impacts of drought on photosynthesis in major food crops and the related mechanisms of plant responses to drought. *Plants* 13 (13), 1808. doi: 10.3390/plants13131808
- Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., and Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* 6, 619. doi: 10.3389/fpls.2015.00619
- Rayaprolu, L., Jayasankar, K., Aarts, M. G. M., and Harbinson, J. (2025). Photosynthetic variation for climate-resilient crops: photosynthetic responses to fluctuating light and chilling in tomato. *Physiol. Plantarum* 177 (3), e70241. doi: 10.1111/pp.70241
- Rosati, A., and Benincasa, P. (2023). Revisiting source versus sink limitations of wheat yield during grain filling. *Agron. J.* 115, 3197–3205. doi: 10.1002/agi.2.21454
- Shi, R., López-Malvar, A., Knoch, D., Tschiersch, H., Heuermann, M. C., Shaaf, S., et al. (2025). Integrating high-throughput phenotyping and genome-wide association analyses to unravel Mediterranean maize resilience to combined drought and high temperatures. *Plant Stress* 17, 1–14. doi: 10.1016/j.stress.2025.100954
- Shin, Y. K., Bhandari, S. R., Jo, J. S., Song, J. W., and Lee, J. G. (2021). Effect of drought stress on chlorophyll fluorescence parameters, phytochemical contents, and antioxidant activities in lettuce seedlings. *Horticulturae* 7 (8), 238. doi: 10.3390/horticulturae7080238
- Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110–124. doi: 10.1016/j.tplants.2015.10.015
- Singh, A., Jones, S., Ganapathysubramanian, B., Sarkar, S., Mueller, D., Sandhu, K., et al. (2021). Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends Plant Sci.* 26, 53–69. doi: 10.1016/j.tplants.2020.07.010
- Song, P., Wang, J., Guo, X., Yang, W., and Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *Crop J.* 9, 633–645. doi: 10.1016/j.cj.2021.03.015
- Stekhoven, D. J., and Bühlmann, P. (2011). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28, 112–118. doi: 10.1093/bioinformatics/btr597
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., and Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Curr. Biol.* 27, R770–R783. doi: 10.1016/j.cub.2017.05.055
- Varshney, R. K., Barmukh, R., Roorkiwal, M., Qi, Y., Kholova, J., Tuberosa, R., et al. (2021). Breeding custom-designed crops for improved drought adaptation. *Adva. Genet.* 2 (3), e202100017. doi: 10.1002/ggn.2.202100017
- Vico, G., Tang, F. H. M., Brunsell, N. A., Crews, T. E., and Katul, G. G. (2023). Photosynthetic capacity, canopy size and rooting depth mediate response to heat and water stress of annual and perennial grain crops. *Agric. For. Meteorol.* 341, 109666. doi: 10.1016/j.agrformet.2023.109666
- Way, D. A., and Yamori, W. (2014). Thermal acclimation of photosynthesis: On the importance of adjusting our definitions and accounting for thermal acclimation of respiration. *Photosynthesis Res.* 119, 89–100. doi: 10.1007/s11120-013-9873-7
- Wen, T., Li, J. H., Wang, Q., Gao, Y. Y., Hao, G. F., and Song, B. A. (2023). Thermal imaging: The digital eye facilitates high-throughput phenotyping traits of plant growth and stress responses. *Sci. Total Environ.* 899, 165626. doi: 10.1016/j.scitotenv.2023.165626

- Xu, R., Ferguson, J. N., Kromdijk, J., and Nikoloski, Z. (2025). Generalizability of machine learning models for plant traits using hyperspectral reflectance data: The case of maize. *bioRxiv*. 2025.07.03.661070, 1–49. doi: 10.1017/9781009157896
- Xu, Y., Zhang, X., Li, H., Zheng, H., Zhang, J., Olsen, M. S., et al. (2022). Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. *Mol. Plant* 15, 1664–1695. doi: 10.1016/j.molp.2022.09.001
- Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214. doi: 10.1016/j.molp.2020.01.008
- Ying, X. (2019). An overview of overfitting and its solutions. *J. Physics: Conf. Ser.* 1168, 22022. doi: 10.1088/1742-6596/1168/2/022022
- Zhang, Z., Qu, Y., Ma, F., Lv, Q., Zhu, X., Guo, G., et al. (2024). Integrating high-throughput phenotyping and genome-wide association studies for enhanced drought resistance and yield prediction in wheat. *New Phytol.* 243, 1758–1775. doi: 10.1111/nph.19942
- Zhou, R., Kan, X., Chen, J., Hua, H., Li, Y., Ren, J., et al. (2019). Drought-induced changes in photosynthetic electron transport in maize probed by prompt fluorescence, delayed fluorescence, P700 and cyclic electron flow signals. *Environ. Exp. Bot.* 158, 51–62. doi: 10.1016/j.envexpbot.2018.11.005