



OPEN ACCESS

EDITED BY

Yuriy L. Orlov,
I.M.Sechenov First Moscow State Medical
University, Russia

REVIEWED BY

Yheni Dwiningsih,
University of Arkansas, United States
Boubacar Gano,
Donald Danforth Plant Science Center,
United States

*CORRESPONDENCE

Viswanathan Chinnusamy
✉ v.chinnusamy@icar.gov.in

RECEIVED 30 April 2023

ACCEPTED 07 June 2023

PUBLISHED 28 June 2023

CITATION

Singh B, Kumar S, Elangovan A, Vasht D,
Arya S, Duc NT, Swami P, Pawar GS,
Raju D, Krishna H, Sathee L, Dalal M,
Sahoo RN and Chinnusamy V (2023)
Phenomics based prediction of plant
biomass and leaf area in wheat using
machine learning approaches.
Front. Plant Sci. 14:1214801.
doi: 10.3389/fpls.2023.1214801

COPYRIGHT

© 2023 Singh, Kumar, Elangovan, Vasht,
Arya, Duc, Swami, Pawar, Raju, Krishna,
Sathee, Dalal, Sahoo and Chinnusamy. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Phenomics based prediction of plant biomass and leaf area in wheat using machine learning approaches

Biswabiplab Singh¹, Sudhir Kumar¹, Allimuthu Elangovan¹,
Devendra Vasht¹, Sunny Arya¹, Nguyen Trung Duc^{1,2},
Pooja Swami¹, Godawari Shivaji Pawar³, Dhandapani Raju¹,
Hari Krishna⁴, Lekshmy Sathee¹, Monika Dalal⁵,
Rabi Narayan Sahoo⁶ and Viswanathan Chinnusamy^{1*}

¹Division of Plant Physiology and Nanaji Deshmukh Plant Phenomics Centre (NDPPC), Indian Council of Agricultural Research (ICAR)-Indian Agricultural Research Institute, New Delhi, India, ²Vietnam National University of Agriculture, Hanoi, Vietnam, ³Division of Agricultural Botany, Vasantrya Naik Marathwada Krishi Vidyapeeth, Parbhani, India, ⁴Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India, ⁵ICAR-National Institute for Plant Biotechnology, New Delhi, India, ⁶Division of Agricultural Physics, ICAR-Indian Agricultural Research Institute, New Delhi, India

Introduction: Phenomics has emerged as important tool to bridge the genotype-phenotype gap. To dissect complex traits such as highly dynamic plant growth, and quantification of its component traits over a different growth phase of plant will immensely help dissect genetic basis of biomass production. Based on RGB images, models have been developed to predict biomass recently. However, it is very challenging to find a model performing stable across experiments. In this study, we recorded RGB and NIR images of wheat germplasm and Recombinant Inbred Lines (RILs) of Raj3765xHD2329, and examined the use of multimodal images from RGB, NIR sensors and machine learning models to predict biomass and leaf area non-invasively.

Results: The image-based traits (i-Traits) containing geometric features, RGB based indices, RGB colour classes and NIR features were categorized into architectural traits and physiological traits. Total 77 i-Traits were selected for prediction of biomass and leaf area consisting of 35 architectural and 42 physiological traits. We have shown that different biomass related traits such as fresh weight, dry weight and shoot area can be predicted accurately from RGB and NIR images using 16 machine learning models. We applied the models on two consecutive years of experiments and found that measurement accuracies were similar suggesting the generalized nature of models. Results showed that all biomass-related traits could be estimated with about 90% accuracy but the performance of model BLASSO was relatively stable and high in all the traits and experiments. The R^2 of BLASSO for fresh weight prediction was 0.96 (both year experiments), for dry weight prediction was 0.90 (Experiment 1) and 0.93 (Experiment 2) and for shoot area prediction 0.96 (Experiment 1) and 0.93 (Experiment 2). Also, the RMSRE of BLASSO for fresh weight prediction was 0.53 (Experiment 1) and 0.24 (Experiment 2), for dry weight prediction was 0.85

(Experiment 1) and 0.25 (Experiment 2) and for shoot area prediction 0.59 (Experiment 1) and 0.53 (Experiment 2).

Discussion: Based on the quantification power analysis of i-Traits, the determinants of biomass accumulation were found which contains both architectural and physiological traits. The best predictor i-Trait for fresh weight and dry weight prediction was Area_SV and for shoot area prediction was projected shoot area. These results will be helpful for identification and genetic basis dissection of major determinants of biomass accumulation and also non-invasive high throughput estimation of plant growth during different phenological stages can identify hitherto uncovered genes for biomass production and its deployment in crop improvement for breaking the yield plateau.

KEYWORDS

high-throughput phenotyping (HTP), RGB image, NIR image, machine learning, i-traits, wheat, shoot area

Introduction

Wheat (*Triticum aestivum* L.) is one of the most important cereal crops in the world since the beginning of agriculture, feeding nearly 40% of the world's population (Giraldo et al., 2019). It is grown in about 217 million hectares across the globe, with an annual production of about around 731 million tonnes (Ramadas et al., 2020). Wheat crop production needs to be increased at least by 60%, to feed the 10-billion people by 2050 (Misra et al., 2020). India is the world's second-largest wheat producer and a major exporter of wheat. Hence wheat crop is given emphasis in crop improvement programs in India as well as in different countries for breeding better wheat varieties with enhanced yield and quality.

Recent advances in the next generation genotyping technologies have helped to cut the cost and time while boosting genotyping precision. At the same time, phenotyping continues to be a barrier in establishing genotype-phenotype relationships (Yang et al., 2020). The introduction and evolution of phenomics in plant science occurred around 2010 with sensors to capture time series information and plant characteristics from digital images, which can ease the phenotyping bottleneck (Yang et al., 2020). Phenomics is the multidisciplinary study of high-throughput accurate acquisition and analysis of multidimensional phenotypes by using digital sensors to capture the morphological and physiological responses of plants (Kumar et al., 2016; Tardieu et al., 2017; Yang et al., 2020). Throughout plant growth and developmental life cycle, phenomics aids in the study of plant morphometry, physiology, leaf color, biomass, seed characteristics, spike number, growth rate and water use efficiency (Chen et al., 2014; Al-Tamimi et al., 2016; Guo et al., 2018; Misra et al., 2020; Elangovan et al., 2023). Researchers across the globe are targeting many plant traits to break the current plateau of the yield (Rauf et al., 2015; Neeraj et al., 2022). The harvest index is one of the most important traits as it links biomass accumulation and grain yield in cereals (Reynolds et al., 2017) and

hence biomass is one of the key traits for crop improvement. Plant leaf area is an important conventional physiological trait used for plant developmental studies and leaf area index estimation (Wu et al., 2022). It is an essential parameter for assessing crop growth and is highly related to the crop biomass and yield (Qiao et al., 2019; Wu et al., 2022).

Manual estimation of plant biomass and shoot area gives accurate information but it is destructive, time and labour intensive, less throughput, and the accuracy is affected due to human errors (Buxbaum et al., 2022) thereby restricting its use in breeding and commercial contexts. Hand-gathered allometric approaches that connect plant volume and height data to biomass are time-consuming, arduous, and may not generalize (Pottier and Jabot, 2017). Also, by the conventional phenotyping, we could only get biomass and leaf area as a single point data, while the high throughput phenotyping empowered generation of time series biomass data (Rahaman et al., 2017; Song et al., 2021). Predicting biomass at multiple stages gives more insight into complex yield architecture (Buxbaum et al., 2022), crop phenotypic (P) and genotypic (G) along with G×E behaviour of plants (Xu, 2016; Van Eeuwijk et al., 2019). As biomass is a time-dependent variable in the plant life cycle, its non-invasive measurement at multiple time points is essential to dissect the complex plant growth characteristics and for its functional mapping (Wu and Lin, 2006; Li and Wu, 2010).

Non-invasive biomass estimation was mainly carried out, in the past, with a single sensor and very few image-based features (Campbell et al., 2015; Al-Tamimi et al., 2016; Rahaman et al., 2017; Asif et al., 2018). Researchers also have tried to predict biomass as linear function of projected area (Golzarjan et al., 2011), multiple linear regression of different parameters, considering both the volume of the plants and their density (Busemeyer et al., 2013; Yang et al., 2014) and with four machine learning based biomass estimation from multiple sensor traits

(Chen et al., 2018b). Many well established machine learning methods have been used earlier for various purposes, such as the prediction of gene expression patterns due to chromatin features (McLeay et al., 2012; Yang et al., 2014; Song et al., 2016), biomass (Chen et al., 2018b), and classification of the disease status of plants (Baranowski et al., 2015). Several models have been developed for in-house experiments to predict biomass of *Arabidopsis* (Arvidsson et al., 2011), barley (Bendig et al., 2015; Chen et al., 2018b), wheat (Golzarian et al., 2011; Parent et al., 2015), and rice (Yang et al., 2014; Campbell et al., 2015), but their reproducibility in other experiments has not been characterized. Also, researchers have tried to predict biomass at early stages (Golzarian et al., 2011; Chen et al., 2018b) which might not cover all the phenotypic variability of plant biomass. Machine learning model-based prediction of biomass and leaf area in wheat under controlled environment condition have not been reported yet.

So, we planned our experiment to develop a generalized robust protocol for non-destructive estimation of biomass and shoot area in wheat at peak vegetative stage by using open-source machine learning tools from the large number of image-based features and from multiple sensors which can be used to precisely predict plant biomass in future experiments by plant scientists. The objectives of this study were (i) to generate multi-experiment phenomics data from multiple sensors to predict plant biomass and shoot area at vegetative stage in wheat (ii) to select the best generic model for accurate prediction of fresh weight (FW), dry weight (DW) and shoot area (SA) by using open source machine learning tools (iii) to identify best surrogate i-Trait for FW, DW and SA. As a result, we screened a generalised model from a large set of machine learning models which considers traits derived from multiple sensors incorporating geometric features, RGB indices, colour class and NIR features covering major determinants of plant growth and also showing higher accuracy across experiments.

Materials and methods

Experimental design for biomass estimation

Two independent experiments were conducted in the Nanaji Deshmukh Plant Phenomics Centre (28°38'31.2"N, 77°09'39.6"E), New Delhi, India, during the winter seasons (Nov-April, Rabi) of 2018 and 2019. In both experiments, wheat germplasm lines and recombinant inbred lines (RILs) of Raj3765 x HD2329 were used. Seeds were sown in pots (0.19 m diameter, 0.4 m high, 15 L volume) containing uniformly filled soil (12.5 Kg per pot). Both experiments differed in genotypes and RILs to validate the model's effectiveness in a wide range of biomass across experiments. Recommended dose of fertilizer (120-80-60 kg/ha N-P-K respectively) was applied to each pot. Well watered condition was maintained in all the pots and recommended weed, pest and disease control practices were followed. Plants were grown in the natural environment for proper growth and biomass accumulation, and shifted to the greenhouse in phase wise manner to capture the biomass at different vegetative stages (before booting). The idea to capture

biomass before booting was due to the fact that after booting the ear creates erroneous result due to its variable weight than leaf and stem. The age of plants were ranging from 30 to 70 days after sowing at the time of imaging. This variability in plant age helped in capturing wider range of FW, DW and SA. Before imaging of the plant, ultra-low weight & solid polypropylene beads were applied just over the soil surface to about 5cm height to arrest direct evaporation from the soil surface, and also for easy segmentation of image. Three hundred plants in 2018 experiment and 154 plants in 2019 experiment were selected for image acquisition using the LemnaTec-Scanalyzer 3D automated phenotyping and imaging platform. Destructive sampling of plants was done to measure above-ground biomass FW (g), DW (g), and SA (cm²). The SA of the whole plant, along with stem and leaf, was measured by using LI-3100C (LI-COR, Lincoln, NE, USA) automatic leaf area meter.

Image acquisition and processing

RGB and NIR images of the plants were taken using a commercial grade RGB (Prosilica GT6600, sensor: ON Semi KAI-29050, LemnaTec, GmbH, Aachen, Germany) and NIR camera (Gold eye P-032 SWIR Cool cameras, sensor: InGaAs, LemnaTec, GmbH, Aachen, Germany) using LemnaTec-Scanalyzer 3D software. Three different side views of RGB images (angles: 0°, 120°, 240°), and one top view RGB image of the plants were captured for each plant using the automated turning and lifting system inside the imaging unit. Three side views were considered, as it is hypothesized that the image from one direction cannot cover all the plant parts; besides, it helps increase the data points corresponding to one plant. NIR sensor captured one side view and one top view image for analysis. A uniform white background was maintained to increase the accuracy of separation between the background and foreground in the images.

Images were processed by the wheat image analysis pipelines developed in the commercial LemnaGrid software. Images were pre-processed to segment the image into foreground and background sections accordingly, and then feature extraction was done to produce a trait list. Extracted traits from the whole dataset were exported in CSV format *via* LemnaGrid and LemnaMiner functionalities, which were used for post-processing and statistical analysis. A detailed data set report is available in (Supplementary Table 1).

Feature analysis and data transformation

After feature extraction, all the features were categorized into four groups: Geometrical features, colour class features, RGB-based indices and near-infrared features. Finally, these features are classified into two major categories namely architectural features (geometric features) and physiological features (colour class, RGB based indices and NIR features). Details of these features are available in Supplementary Table 1. These features were specified by considering the type of imaging sensors (RGB and NIR) and object orientations (side and top views). All the traits were curated for redundancy, processing error, outliers, and non-informativeness

by both statistical approach (Multicollinearity removal, poor heritability, etc.) and manual curation. We kept as much variation and informative features as possible to improve the model accuracy for biomass estimation.

Each experiment dataset was transformed into matrix $X_n \times m$ where “n” is the number of plants and “m” is the number of phenotypic traits. Plants represented rows, and different traits represented columns. All missing value plants were discarded for reduction of data analysis complicacy. Before applying regression models, all datasets were normalized as described by (Chen et al., 2018a).

Phenotypic data interpretation and visualization

A phenotypic similarity tree was used to see the correlation between all the traits and the similarity between experiments. Principle component analysis was performed on the transformed data matrix $X_n \times m$ in the same way as described by (Chen et al., 2014) for all the experiments. Both correlation and PCA analysis was done in R software (R Core Team, 2021). All visualization graphs were produced using “ggplot2” package in R software (R Core Team, 2021).

Modelling for predicting plant biomass

After i-Trait selection and phenotypic analysis, the next part was to fit the selected data into the model to predict FW, DW, and SA. We used the open-source tool “HTPmod” for modelling (Chen et al., 2018a). In HTPmod (Shiny framework-based application), the module *predMod* contains 16 models constructed with 16 different machine learning methods to regress input features to output traits of interest. The description and details of all the models is available in [supplementary data sheet 1](#). We used the default hyperparameters applicable for different machine learning models present in the HTPmod application. All the model parameters were controlled using respective R package (Given in [supplementary data sheet 1](#)). Also, for additional tuning functionality of “caret” R package was used.

Evaluation of the models performance

Model performance was evaluated using k fold cross-validation method and N-times randomization, where we assigned k and N to 10. So, we adopted a 10-fold cross-validation strategy and ten times randomization for model evaluation by considering the average value. The data set was randomly divided into a training set of 90% of plants and a testing set of the remaining 10% of plants. Then each model was run to predict FW, DW, and SA for the testing data, and then the predicted biomass was compared with the manually measured FW, DW, and SA.

All the regression models were evaluated by the Pearson correlation coefficient (PCC; r), the coefficient of determination

(R^2) and the root mean squared relative error of cross-validation (RMSRE) between the predicted and observed values (Chen et al., 2018a).

Results

i-Trait extraction and characterization

We analysed two image datasets, 1800 images (Experiment 1, 2018) and 924 images (Experiment 2), collected from 300 (Experiment 1) and 154 plants (Experiment 2, 2019), respectively. An overview of the experimental site and experiment is shown in [Figure 1](#). Each plant was imaged at a single time point by RGB (One top view and three side view images at 0°, 120° and 240° angle) and NIR sensor (One top view and one side view), then plants were harvested to measure FW) and SA immediately, and oven dried to record DW. To increase the variability in biomass range, we conducted our experiment with germplasm and RILs with wider variability in biomass, and phenotyping at different plant growth stages. All the images retrieved from the server and processed by the LemnaGrid image analysis pipeline, which was modified explicitly for mid to large-sized important cereals such as wheat (challenging due to its narrow leaf and compact character), resulting in nearly 200 phenotypic traits extracted from images of each plant ([Figures 2A, B](#)). After quality control such as outliers, multicollinearity and manual checking of all the extracted traits, we selected 77 i-Traits ([Figure 2C](#)) which were divided into two major categories of traits such as architectural (35 traits) and physiological (42) traits ([Supplementary Table 1](#)).

Phenotypic profile analysis of plants in both experiments

We observed broader range of phenotypic values in all the traits in both experiments. The phenotypic value of FW, DW, and SA of experiment 1 ranged 0.6-145.86 g, 0.15-27.35g and 18.43-3622.25 cm², respectively. In the experiment 2, the FW, DW, and SA ranged 3.3-107.8 g, 0.5-22.2 g, and 26.15- 1434.01 cm², respectively ([Figures 3C-E](#)).

In both experiments, 77 selected i-Traits were analysed. Principal component analysis (PCA) was carried out for both experiments to see the global phenotypic variation present in the population. The top 4 principal components (PCs) of experiment 1 and 2 accounted for 76.56% and 75.02% of the total phenotypic variation explained by i-Traits. The first two PCs clearly distinguished the experiments as the first two PCs of experiment 1 and 2 accounted for 39.41%, 17.77% and, 29.03%, 23.28%, respectively ([Figures 3A, B](#)).

To access the patterns of trait correlations, we performed the trait similarity analysis based on canonical Pearson’s correlation coefficient ([Figures 4A, B](#)). We observed that the patterns of correlation were similar in both experiments however i-Traits were more correlated in experiment 1 than in experiment 2 as suggested by the intensity of colour in [Figures 4A, B](#) but traits across



FIGURE 1

Pictorial representation of experimental site and setup. (A) Nanaji Deshmukh Plant Phenomics Center (NDPPC), Indian Council of Agricultural Research–Indian Agricultural Research Institute, New Delhi, India. (B) Four climate controlled green houses present within the facility. (C) Ongoing wheat plant experiment for wheat non invasive biomass prediction. (D) LemnaTec system controller computer for image acquisition and processing. (E) RGB, side view image of wheat plant inside imaging chamber. (F) Destructive sampling for biomass measurement.

experiments shows similar correlation with each other. Also, the correlation of architectural traits with physiological traits was less, and traits were both positively and negatively correlated in the two experiments. Correlation among the physiological traits as higher than the correlation among the architectural traits.

Phenotypic association of i-Traits with FW, DW and shoot area

We further evaluated the association of i-Traits with the manual traits in both experiments to see the relation between i-Traits and manual traits and to observe the biological relevance of predicting

biomass from digital image-derived parameters (Figure 5). The direct relatedness of architectural traits with biomass is well reported in literature and the physiological traits that we included here having some unique feature that have not been reported earlier such as RGB colour class and RGB indices. The use of physiological traits in biomass estimation is to incorporate the additional properties presented by NIR grey value (water status of plant), RGB colour class (greenness of different plant pixels) and RGB indices (reported vegetation indices). The PCC of the i-Traits with FW, DW and SA in experiment 1 ranged from -0.73 to 0.97, -0.76 to 0.92, and -0.76 to 0.93, respectively. While in the case of experiment 2, the FW, DW and SA correlations with i-Traits ranged from -0.76 to 0.95, -0.74 to 0.93 and -0.73 to 0.95, respectively. Architectural traits had a higher

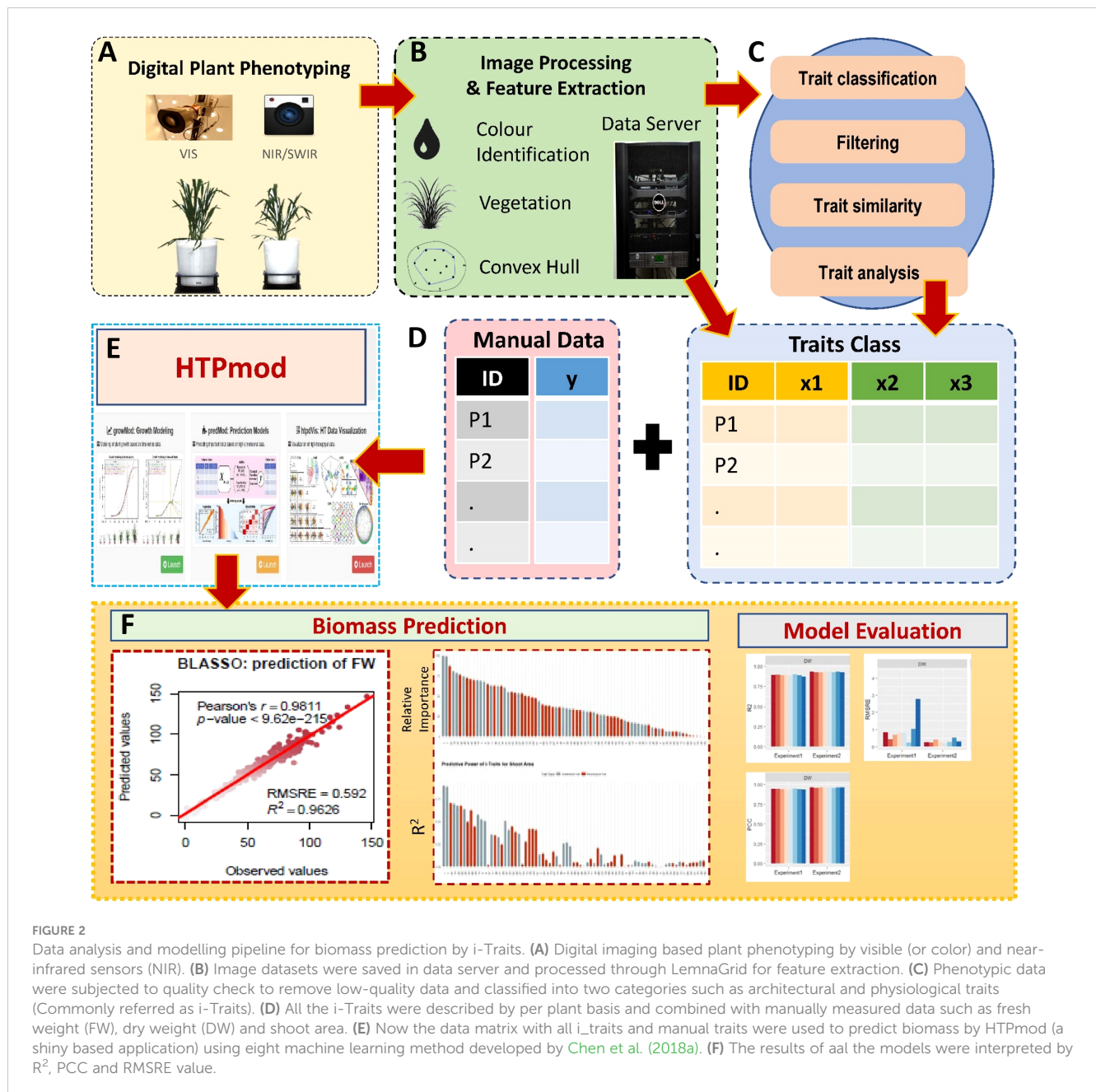


FIGURE 2

Data analysis and modelling pipeline for biomass prediction by i-Traits. (A) Digital imaging based plant phenotyping by visible (or color) and near-infrared sensors (NIR). (B) Image datasets were saved in data server and processed through LemnaGrid for feature extraction. (C) Phenotypic data were subjected to quality check to remove low-quality data and classified into two categories such as architectural and physiological traits (Commonly referred as i-Traits). (D) All the i-Traits were described by per plant basis and combined with manually measured data such as fresh weight (FW), dry weight (DW) and shoot area. (E) Now the data matrix with all i_traits and manual traits were used to predict biomass by HTPmod (a shiny based application) using eight machine learning method developed by Chen et al. (2018a). (F) The results of all the models were interpreted by R^2 , PCC and RMSRE value.

correlation with manual traits, than physiological traits. Among all the traits, PSA and Area_SV had a correlation of >0.9 for FW, DW and SA in both experiments. As expected, both boundary point count and compactness of side view images are also highly correlated with the manual traits. This explains that the biologically relevant traits, such as area and architecture, are highly related to biomass traits. We also found that physiological traits such as colour class correlated with biomass traits. NIR_SV was negatively correlated with FW, DW and SA ranging from -0.36 to -0.68 in both experiments. This also suggests that physiological traits not directly measured as plant architectural traits can also be used as biomass predictors. Indices derived from the mean blue, green and red values of RGB images also correlated with biomass, but the correlations were relatively lower than other traits.

Modelling of plant biomass using machine learning methods

HTP is particularly meaningful in dissecting complex genetics of biomass development in plants. The relationship between i-traits and above-ground shoot biomass accumulation were analysed by digital phenotyping data containing structural (e.g., architectural traits) and physiological traits (e.g., colour class, RGB based indices and plant moisture content as reflected by NIR-intensity traits). The results about i-Traits and manual traits suggested that these i-Traits can be very useful in predicting plant biomass-related traits.

To develop the model by machine learning (ML) methods, we used the available open-source tools. We used the *predMod* module from the HTPmod, a R program based shiny application. We used

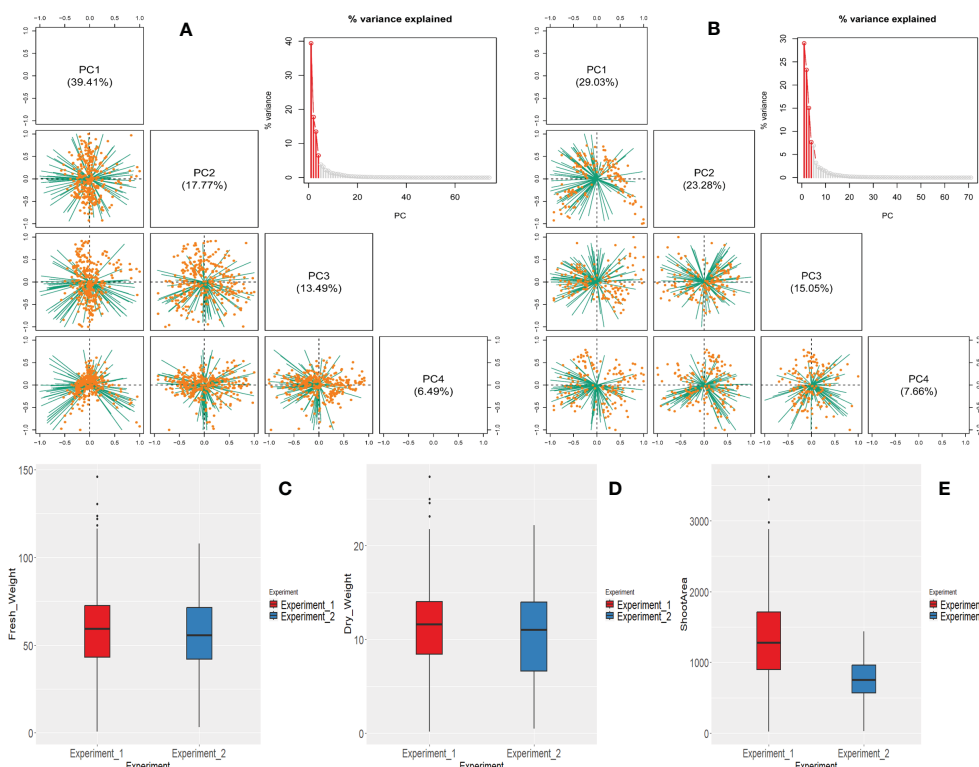


FIGURE 3

Characterization of all the i-Traits by phenotypic analysis for both experiments (Experiment 1 and Experiment 2). (A, B) Principal component analysis of all the i-Traits for experiment 1 and experiment 2 respectively. Four PCs represented here which captured approximately all the variation in the data. Variance proportion explained by the PCs is shown in parentheses. Individuals are represented by orange dots and traits are represented by green lines. (C–E) Differences in the FW, DW and shoot area across experiments represented by box plot analysis.

16 ML methods available in *predMod*. From the 16 models, we found eight models that were consistent for all the manual trait estimations in both the experiments. The selected models were Bayesian LASSO, Bayesian regularized neural networks (BRNN), Lasso and elastic-net regularized generalized linear models (GLMNET), Gaussian process with the polynomial kernel (GP-Poly), multivariate adaptive regression splines (MARS), random forest (RF), ridge regression (RIDGE) and Support vector machines with linear kernel (SVM-Linear).

Biomass estimation model performance evaluation

Since our aim was to produce a more generalized model that can use genotypes and RILs to predict the biomass from image traits over multiple time points, the models were tested in both experiments to validate their performance and to evaluate their generalized nature. Analysis was performed with all 16 models, from which eight models that performed better for all the traits irrespective of experiments were selected. Results of all the 16 models showed that the R^2 for FW prediction was between 0.84 to 0.96, DW prediction was between 0.79 to 0.90 and SA prediction was 0.85 to 0.97 in experiment 1 (Supplementary Data sheet 1). In experiment 2 the R^2 for FW prediction was between 0.88 to 0.96, DW prediction was between 0.86 to 0.93 and SA prediction was 0.86

to 0.93 (Supplementary Data sheet 1). We found that eight models performed relatively better than other eight models. Bayesian Generalized Linear Model (BGLM), Gradient Boosting Machine (GBM), Generalized Linear Model (GLM), Gaussian Process with Radial Kernel (GP-Radial), K-Nearest Neighbors (KNN), Least Absolute Shrinkage and Selection Operator Regression (LASSO), Multivariate Linear Regression (MLR) and Support Vector Machines with Radial Kernel (SVM-Radial) performed with relatively less accuracy than other eight models namely Bayesian LASSO, Bayesian regularized neural networks (BRNN), Lasso and elastic-net regularized generalized linear models (GLMNET), Gaussian process with the polynomial kernel (GP-Poly), multivariate adaptive regression splines (MARS), random forest (RF), ridge regression (RIDGE) and Support vector machines with linear kernel (SVM-Linear). BGLM, GBM, GLM, GP-Radial, KNN, LASSO, MLR and SVM-Radial had R^2 value for prediction of FW ranged from 0.84 to 0.90, for DW ranged from 0.79 to 0.87 and for SA ranged from 0.85 to 0.95 in experiment 1 (Supplementary data sheet 1). At the same time BLASSO, BRNN, GLMNET, GP-Poly, MARS, RF, RIDGE and SVM-Linear had higher R^2 value of 0.94 to 0.96 for FW, 0.87 to 0.90 for DW and 0.93 to 0.96 for SA prediction in experiment 1 (Figures 6–8). Similar condition was there in experiment 2 also. In experiment 2, BGLM, GBM, GLM, GP-Radial, KNN, LASSO, MLR and SVM-Radial had R^2 value of 0.88 to 0.92 for FW, 0.86 to 0.89 for DW and 0.86 to 0.90 for SA prediction (Supplementary data sheet 1). While BLASSO, BRNN,

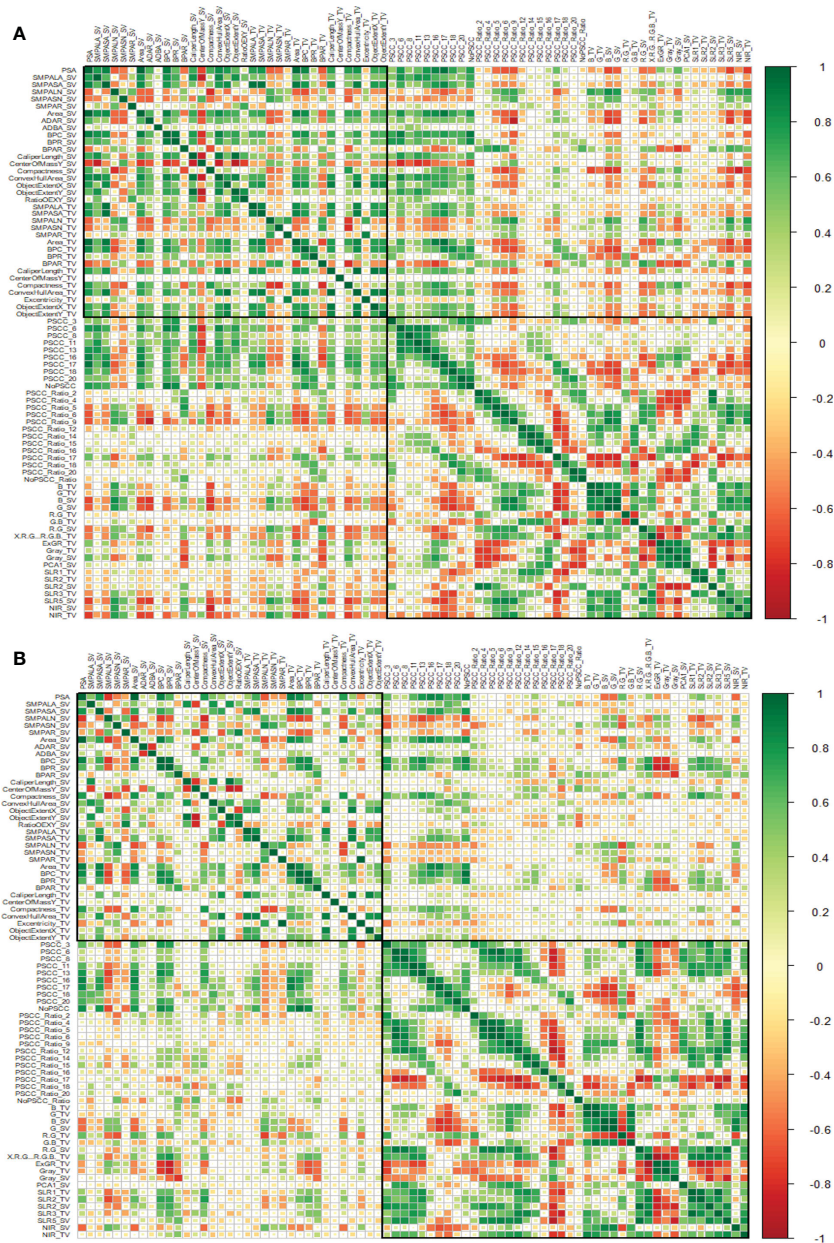


FIGURE 4 Assessment of trait similarity between i-Traits across experiments. **(A)** Canonical correlation analysis of i-Traits based on experiment 1 (Top) and **(B)** experiment 2 (Bottom). Heatmap plot is organized by both architectural and physiological traits represented by two highlighted boxes. Top box represents Architectural traits and bottom box represents physiological traits.

GLMNET, GP-Poly, MARS, RF, RIDGE and SVM-Linear had higher R^2 value of 0.94 to 0.96 for FW, 0.91 to 0.93 for DW and 0.90 to 0.93 for SA prediction in experiment 2 (Figures 6–8). We also noticed that these eight models performed better than normal multivariate regression (MLR) model. All eight selected models BLASSO, BRNN, GLMNET, GP-Poly, MARS, RF, RIDGE and SVM-Linear performed relatively similar in terms of estimation accuracy, while we confined the results to select four best-suited models for our experiments namely BLASSO, BRNN, GLMNET and GP-Poly that were with less RMSRE value and with similar estimation accuracy across experiments.

In case of FW estimation, R^2 and r value ranged from 0.95 to 0.96 and 0.97 to 0.98 in the two experiments for BLASSO, BRNN, GLMNET and GP-Poly models (Figure 6). So, this represents the generalized nature of the estimation, as almost all four models showed similar results. DW's estimation also showed promising results as the R^2 and r value ranged from 0.89 to 0.93 and 0.94 to 0.96 in the two experiments (Figure 7). In SA estimation, R^2 and r value ranged from 0.91 to 0.97 and 0.95 to 0.98 in both the experiments (Figure 8). As per these four models, Experiment 1 had slightly higher estimation accuracy than experiment 2 for FW and SA estimation but in the case of the DW, experiment 2 had

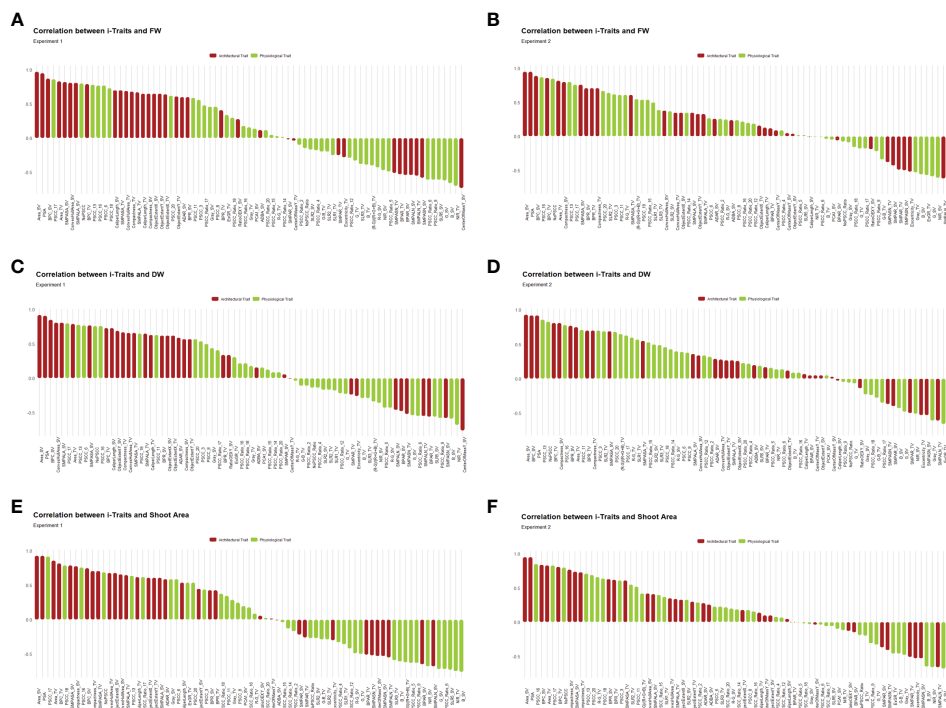


FIGURE 5

Pearson's correlation coefficients (PCC) in each experiment were calculated to assess the relationship between i-Traits and manual traits. The PCC were consistent in both experiments and most of the features were having more than 0.5 positive or negative correlation coefficients with FW, DW and shoot area. (A, B) PCC between FW and i-Traits in experiment 1 and 2 respectively. (C, D) PCC between DW and i-Traits in experiment 1 and 2 respectively. (E, F) PCC between shoot area and i-Traits in experiment 1 and 2 respectively.

better performance than experiment 1. RMSRE values were relatively lower in experiment 2 than in experiment 1 in most of the parameters but the difference was not large.

By observing the estimation ability of the models by R^2 and r and RMSRE we found out BLASSO, BRNN, GLMNET and GP-Poly were relatively better-performing models in all the traits and across all the experiments. Selecting a single model was interesting due to the relatively similar and better estimation accuracy across models. Nevertheless, BLASSO performed better in all the trait estimations in all the experiments with the highest R^2 and lowest RMSRE values. The R^2 of Bayesian Least Absolute Shrinkage and Selection Operator BLASSO for FW prediction was 0.96 (both year experiments), for DW prediction was 0.90 (Experiment 1) and 0.93 (Experiment 2) and for SA prediction 0.96 (Experiment 1) and 0.93 (Experiment 2). Also, the RMSRE of BLASSO for FW prediction was 0.53 (Experiment 1) and 0.24 (Experiment 2), for DW prediction was 0.85 (Experiment 1) and 0.25 (Experiment 2) and for SA prediction 0.59 (Experiment 1) and 0.53 (Experiment 2).

Evaluation of i-Traits for predictive power and relative importance

For each i-Trait, the predictive power and relative importance were calculated by the degenerate model using the *predMod* module. In both experiments, the predictive power and feature importance of i-Traits were similar, so here we explained the

individual capability of each i-Trait as a predictor of biomass in experiment 1 (Figure 9) and experiment 2 (Supplementary data sheet 1). Both predictive power and relative importance of most architectural traits were higher than physiological traits. The top 10 important features contained architectural and physiological traits, but more architectural traits were present than physiological ones. PSA, Area_SV, and BPC_SV have higher predictive power and relative importance among all the i-Traits for FW, DW and SA. Area_SV had R^2 of 0.94 for the estimation of FW, which was highest than the DW (R^2 of 0.79) and shot area (R^2 of 0.89) estimation. Among physiological traits, PSCC_17, 16, 13, NIR_TV and SV, B_SV, G_SV, SLR5_SV, Gray_TV, etc., had higher predictive power and relative importance than other physiological traits. NIR_TV and NIR_SV also had significant predictive power and relative importance ranging from ~0.1 to 0.5 (R^2) and ~5 to 59% (Inclusive MSE) relative importance. The best predictor i-Trait for FW and DW prediction was Area_SV and for SA prediction was projected shoot area. These results give deep understanding into major determinants of plant biomass and also suggests that along with architectural traits, physiological traits also help to improve estimation accuracy and are determinants of plant growth.

Discussion

Aided by multiple imaging sensors and computer vision optimization by machine learning methods, high throughput

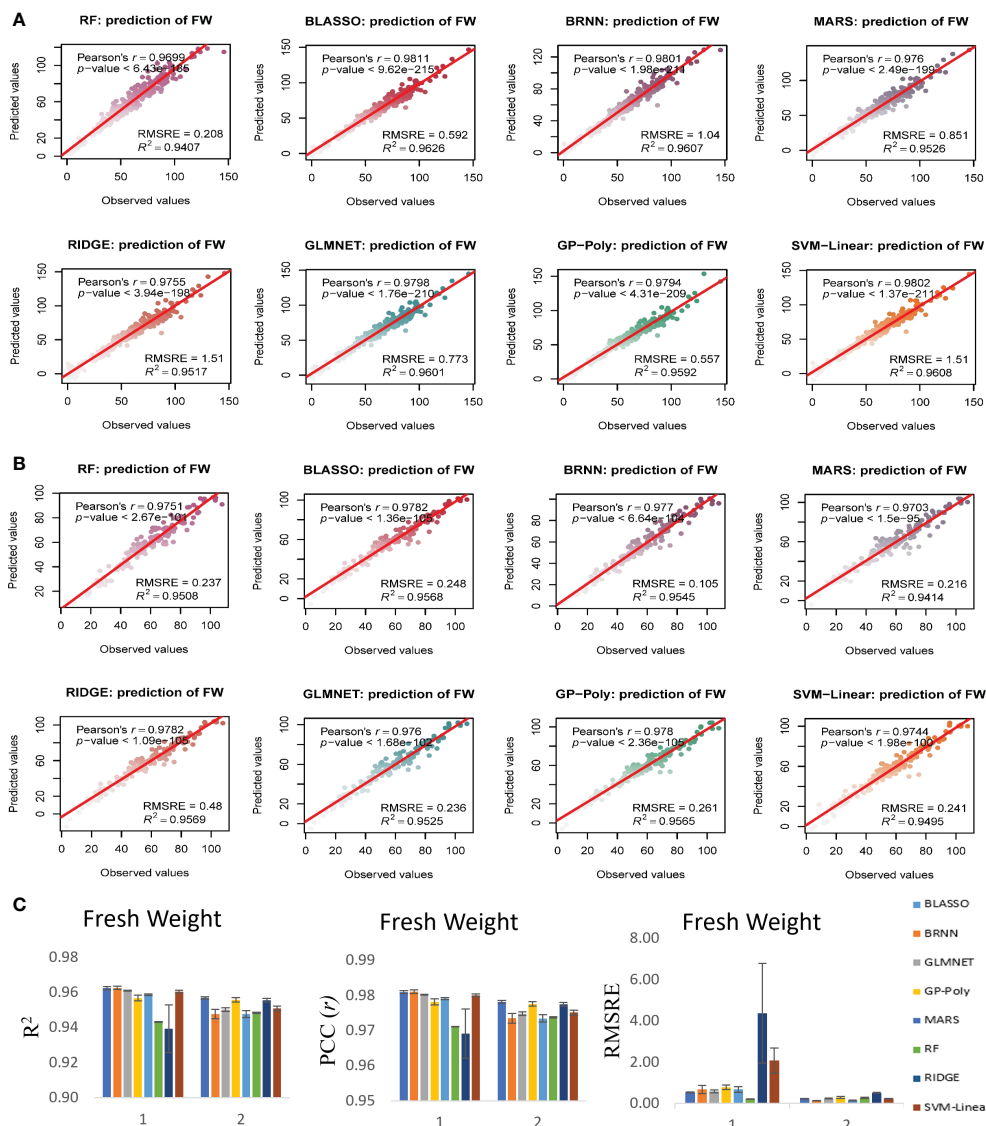


FIGURE 6

Non-destructive estimation of above ground plant biomass (AGPB) with i-Traits, using *predmod* in HTPmod. Scatter plots of observed FW vs predicted FW values using 8 prediction models based on machine learning methods BLASSO, BRNN, GLMNET, GP-Poly, MARS, RF, RIDGE and SVR. The prediction models were evaluated by Pearson's correlation coefficient (r), its corresponding P -value, R^2 and RMSRE. (A, B) Prediction of FW in experiment 1 and 2 respectively. (C) Evaluation of the model performance of each regression model used for AGPB prediction.

phenotyping has emerged as a major technology helping to dissect plant biological characters unseen before (Kumar et al., 2016; Singh et al., 2016; Strock et al., 2022). So, gathering multi-dimensional data over multiple time points at multiple organ levels is the critical component of HTP, which has boosted crop improvement programs (Roth et al., 2021; Pérez-Valencia et al., 2022). Harvest index has been a critical trait for cereal breeders for decades in crop improvement programs, which is closely related to the ability of the plant to accumulate plant biomass and convert it into yield (Chao et al., 2019; Porker et al., 2020). Since, already we have reached a plateau in improving harvest index, for further improvement in yield, we must improve biomass. We need to characterize accurate biomass of plants at different stages to identify genes working at different phenological stages for biomass production (Chang et al., 2019; Chen et al., 2020; Rabab et al., 2021). Also, data at multiple

time points are needed for the functional mapping of plant biomass accumulation. Functional mapping can be a powerful tool to reveal transient and deterministic quantitative trait loci (QTLs) for a complex trait such as biomass (Wu & Lin, 2006; Camargo et al., 2018; Jiang et al., 2019; Lyra et al., 2020). Plant leaf area is a critical physiological parameter which determine the ability of plant to produce economic yield and also it helps to determine the plant leaf area index which is an important conventional physiological trait used for plant developmental studies (Wu et al., 2022). Leaf area index (LAI), is a critical parameter of wheat growth, can provide dynamic information during wheat growth phases and closely associated with crop biomass and yield. So, shoot area estimation by non-destructive methods will definitely help to the plant science community in the future. As per our knowledge no work has been done particularly for image-based shoot area estimation by

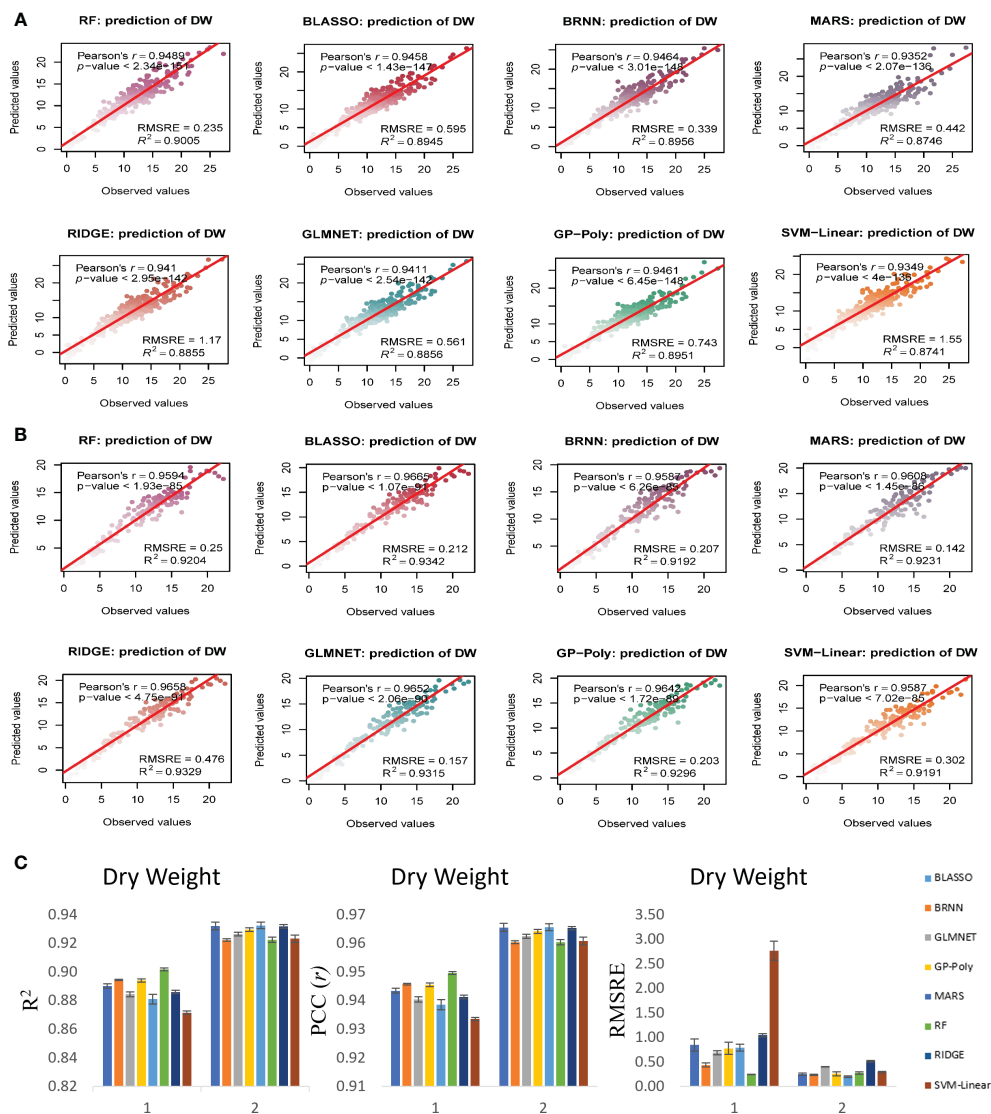


FIGURE 7

Non-destructive estimation of DW by using i-Traits. (A, B) Prediction of DW in experiment 1 and 2 respectively. (C) Summary of the predictive power of each regression model.

validating with actual shoot area in wheat as it has a compact plant architecture. Traditionally, biomass measurement has been done by destructive methods, which is laborious and time-consuming. Also, one of the other major lacuna was the inability to take data points at multiple time scales (Tackenberg, 2007; Buxbaum et al., 2022). Therefore, to address this major bottleneck, automated, non-destructive biomass estimation by digital imaging method is gaining importance since the evolvement of high throughput phenotyping (Golzarian et al., 2011; Rahaman et al., 2017; Chen et al., 2018b; Buxbaum et al., 2022).

Several studies in the recent past have developed different models for biomass estimation, but their applicability in other in-house experiments has not been tested widely. Golzarian et al. (2011) and Chen et al. (2018b) predicted biomass in wheat at the age of 15 to 43 days after sowing and in barley at the age of 27 to 58 days after sowing respectively, where crop biomass is not very high and overlapping of leaves are less. But it is challenging to predict the

biomass of wheat plant at peak vegetative growth. Also, different crops pose different plant architecture such as leaf distribution, tiller numbers overlapping of leaves etc which affects the biomass prediction model accuracy. So, a robust and open-source model that excludes the need for repeated destructive measurement is the need of the hour for any high throughput phenotyping facility worldwide. To address this, we conducted an experiment to estimate wheat biomass and related traits non-destructively in the largest phenomics facility in India, Nanaji Deshmukh Plant Phenomics Centre (NDPPC). We aimed to develop a pipeline to predict plant FW, DW and SA non-destructively by open-source tools, which can be used in future experiments by different researchers. We conducted two experiments in consecutive years to see if the models work equally well in different datasets.

The selection of predictor variables for the estimation of a trait is necessary. In the past, single i-trait based biomass estimation (Tackenberg, 2007; Golzarian et al., 2011; Campbell et al., 2015) and

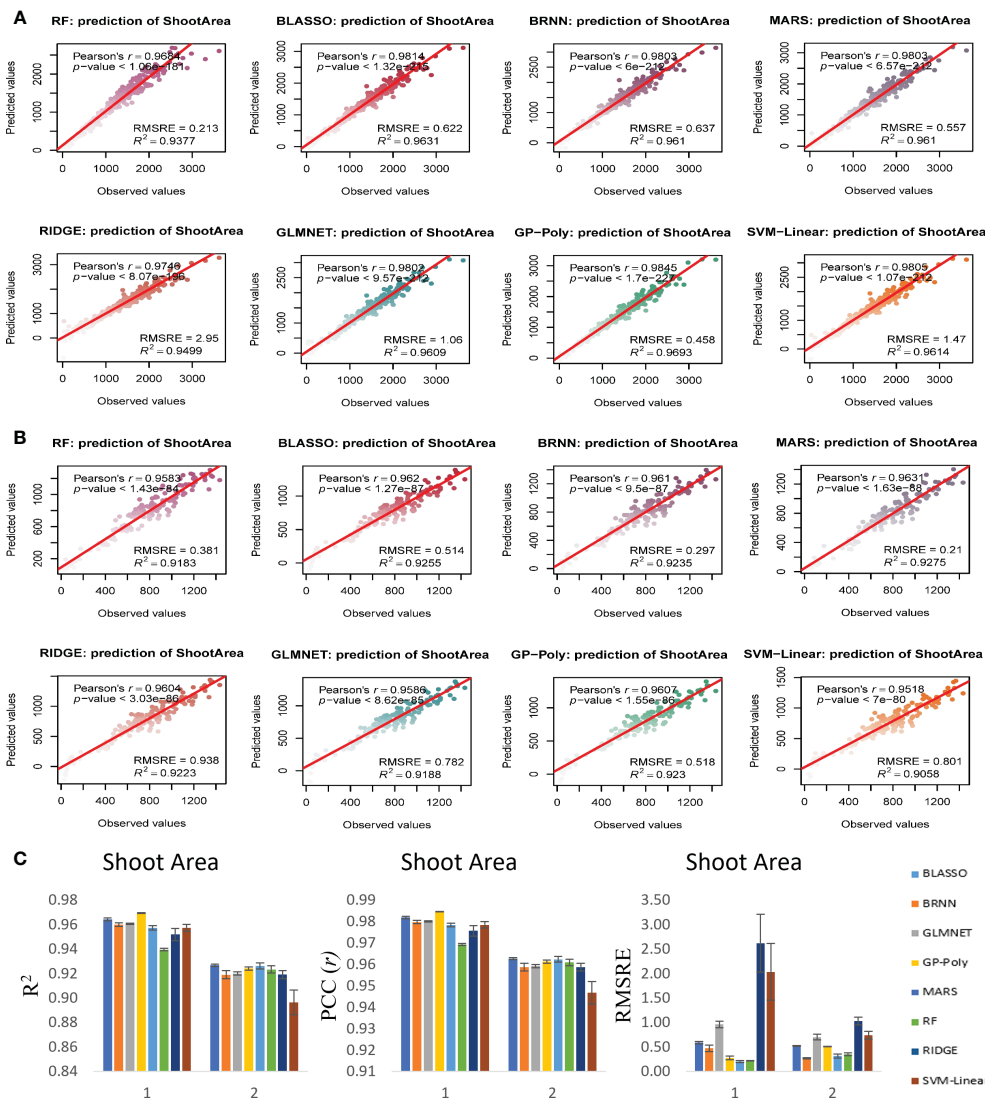


FIGURE 8 Non destructive estimation of shoot area by using i-Traits. (A, B) Prediction of shoot area in experiment 1 and 2 respectively. (C) Summary of the predictive power of each regression model.

modified experiments with multiple traits have been carried out (Yang et al., 2014; Rahaman et al., 2017; Chen et al., 2018b). Previous studies have suggested that combining multiple traits such as vegetation indices and plant-height information can improve biomass estimates (Bendig et al., 2015; Han et al., 2019). Reports have shown that multiple traits and multi-sensor based estimation of biomass have more accuracy and biological meaning (Chen et al., 2018b) which includes different categorical features such as geometric or architectural traits, colour based and NIR based physiological traits. NIR reflectance have been reported to determine water status of plants (Neilson et al., 2015; Jin et al., 2017). As two plants having same plant architecture but different water status will differ in their fresh weight due to the differences in water status. By taking a leaf out of the literature review of non-destructive plant biomass, we selected both RGB and NIR-based traits for our study. We included architectural, colour class based, mean red, blue and green-based indices and mean grey values to

improve the estimation accuracy and make a robust model. We selected 77 high-quality i-Traits by removing redundant, non-informative traits using statistical and manual methods. Both experiments 1 and 2 have differed in phenotypic responses as accessed by principal component and box plot analysis (Figure 3) which emphasizes the independent nature of experiments. So, this will help in understanding the generalised nature of the models across experiments.

We predicted FW, DW and SA using 77 i-Traits by 16 machine learning methods which was available in open-source tool “HTPmod”. To validate the results, we also ran all the models in another dataset (experiment 2) with the same i-Traits for the estimation of FW, DW and SA. The estimation accuracy was consistent in both experiments which suggests that all those models were generalized in nature and selecting a particular model will not discriminate much accuracy. By observing the predictive power, PCC and RMSRE, we found eight models to be

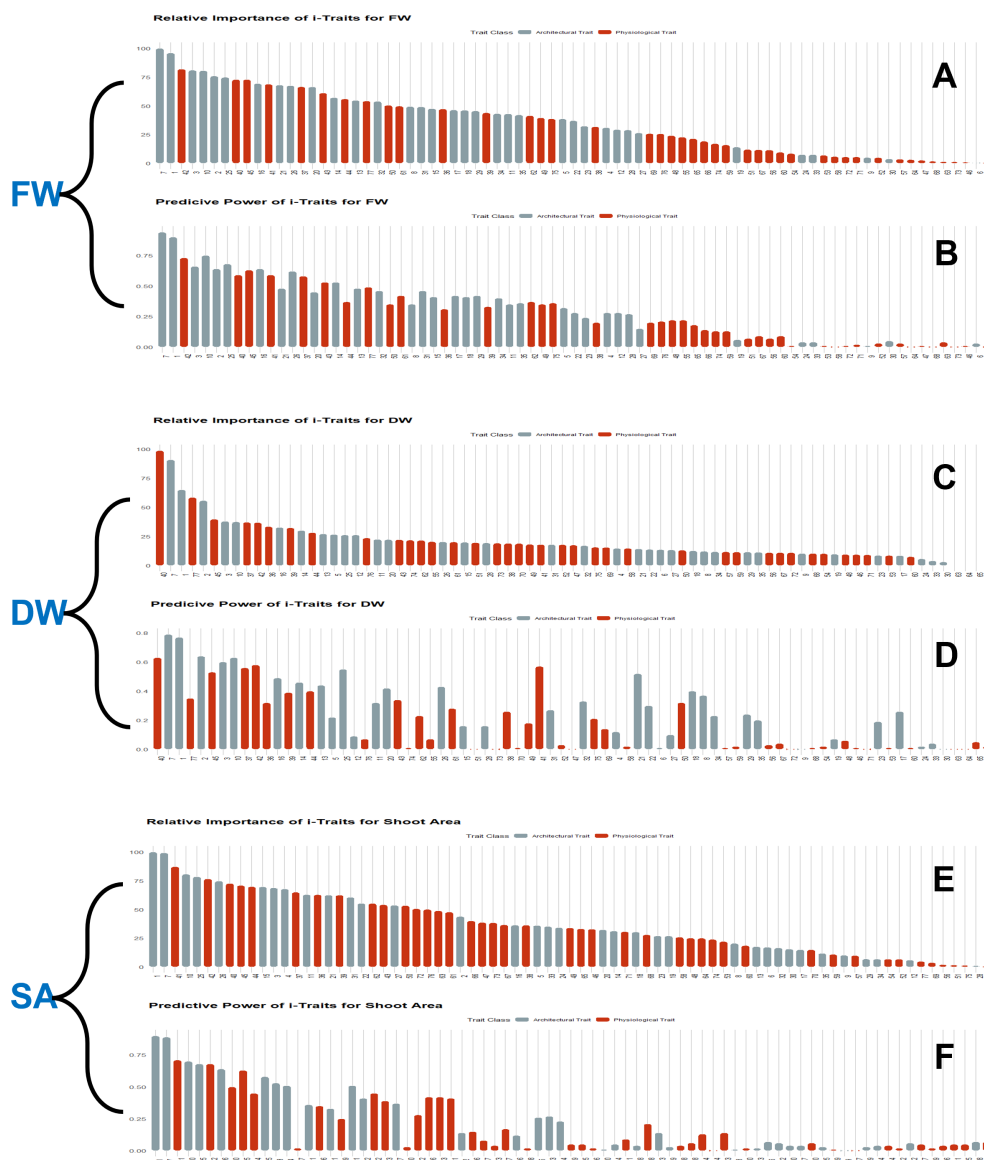


FIGURE 9 Estimation of predictive power and relative importance of features (i-Traits) used in regression models. Bayesian-LASSO (BLASSO) model was used for wheat FW, DW and shoot area prediction, using i-Traits from two independent experiments. But as both experiments performed relatively similar with respect to prediction power and relative importance, experiment 1 information is presented here for the ease of understanding. (A, C, E) Relative importance of i-Traits used for FW, DW and shoot area prediction respectively. (B, D, F) Predictive power of i-Traits used for FW, DW and shoot area prediction respectively.

performing relatively similar with higher estimation accuracies. We included multivariate adaptive regression splines (MARS), random forest (RF) and support vector machine-Radial regression (SVR) in our selected eight models as they had accurate biomass predicting ability (Chen et al., 2018b). All the predictive models worked accurately for FW, DW and SA in both experiments. But we found that eight models namely BGLM, GBM, GLM, GP-Radial, KNN, LASSO, MLR and SVM-Radial performed relatively poor than other eight models namely BLASSO, BRNN, GLMNET, GP-Poly, MARS, RF, RIDGE and SVM-Linear. Here it is interesting to see that simple multivariate linear regression was in the relatively poor performing group which suggests that use of complex machine learning model improved the accuracy of prediction than simple

MLR model (Chen et al., 2018b). The main reasons for the difference in the prediction accuracy of different machine learning models depends upon factors such as model assumption parameters, model architecture, overfitting of model, sensitivities to extreme values, collinearity present within the data set, complexity of the model and number of samples (Ogutlu et al., 2012; Arruda et al., 2015; Esposito et al., 2019; Adak et al., 2021). From all the 16 models four models performed relatively poor than other models namely BGLM, GP-Radial, KNN and SVM-Radial. BGLM model is less suited for small scale data and affected by multicollinearity (Koehrsen, 2018), GP-Radial has problem of scalability, representational power and targeted optimisation (Hensman et al., 2013; Krauth et al., 2016), KNN mostly used on

classification and detection based problems and has limitations of poor performance under high dimensionality and sensitive to extreme values (Uddin et al., 2022) and SVM-Radial has sensitivities to extreme values, collinearity and more helpful for classification problems. Although these are very good machine learning algorithms cited in literature and worked well under different datasets but in our dataset and type of data, they performance were hampered.

However, for general use and model selection purposes, we made the more stringent model selection with respect to robustness, predictive power, PCC and RMSRE for ease of use in future experiments. The four models BLASSO, BRNN, GLMNET and GP-Poly were consistent across the experiments in FW, DW and SA estimation with relatively similar predictive power, PCC and RMSRE value. These four models had higher R^2 and PCC value and lower RMSRE value than other models. The estimation accuracies of these models were at par with the accuracy that have been reported (Neumann et al., 2015; Rahaman et al., 2017; Chen et al., 2018b) in literature and also the reproducibility of the accuracy was validated by the results from our second experiment. Also, the validation of actual leaf area to predicted SA has been a unique feature of this undertaken research. Among all these models we found BLASSO to be the best model in terms of high accuracy and low RMSRE value for all the biomass and related trait prediction and was also stable and interpretable across experiments. BLASSO had highest R^2 value and PCC value for both year experiment and lowest RMSRE value among all models. BLASSO is a popular high dimensional data analysis method. It can perform regularization and variable selection at the same time. This can increase the precision of predictions and interpretation of a problem (Vasquez et al., 2016). It has major advances in terms of assumptions regarding the sample distribution as it is independent of normality of sample distribution, very efficient in handling large data set and dimensionality also helps to overcome underfitting of model data. So, with the help of this model along with multi-sensor dataset, we were able to predict FW, DW and SA accurately in wheat.

Our results showed that by using open-source tools, we can predict FW, DW and SA of wheat plant accurately in any in-house experiment. The estimation accuracy of all the traits, such as FW, DW, and SA, across both experiments was consistent with past studies (Golzarian et al., 2011; Campbell et al., 2015; Neilson et al., 2015; Neumann et al., 2015; Rahaman et al., 2017; Chen et al., 2018b). It also reflected how individual trait impact estimation by analysis of the relative importance and predictive power of each trait. The critical features such as PSA, Area_SV, BPC_SV, PSCC_17, 16, 13, NIR_TV, SV, B_SV, G_SV, SLR5_SV and Gray_TV, which had consistently higher predictive power and more significant correlation with FW, DW and SA can be used as a surrogate for biomass accumulation and dissection of their genetic basis in crop improvement programs. As biomass is a complex trait showing both spatio-temporal variations upon different environments, its characterization can be done using the above-said i-Traits and models effectively at multiple time points.

There was no major limitation of this study but found some interesting challenges which can be addressed by conducting further research. This study was conducted in the controlled

greenhouse conditions so the models selected here have to be tested to predict the biomass under field condition to check the generalised nature of the models. Increase in number of samples and testing of models in different abiotic stress conditions will give deeper insights into biomass modelling. As we have witnessed increase in prediction accuracy was achieved by multiple sensor data, use of advanced hyperspectral sensor data and LIDAR sensor data in further research can give new insight into prediction accuracy of machine learning models.

Conclusion

We developed the biomass and leaf area estimation model using the Bayesian Least Absolute Shrinkage and Selection Operator (BLASSO) machine learning method with high accuracy, which will be helpful to future researchers in predicting biomass and leaf area with high accuracy and robustness. We used the broad genotypic base to include all possible variations in the biomass estimation model to make it more robust. In order to bring out novelty in our research, we used 16 machine-learning models to identify the best estimation model. Wide range of phenotypic variations were taken into consideration by mixing genotypes and recombinant inbred lines (RILs) at different phenological stages. We predicted SA with higher accuracy in wheat crop having compact plant architecture which had not been validated earlier. We considered consecutive-year data set to evaluate the model replicability. We found Bayesian Least Absolute Shrinkage and Selection Operator (BLASSO) to be the best model which gives prediction accuracy of 0.96 for FW, 0.90 for DW and 0.96 for SA. So, this model can be used as a generic model to predict the vegetative stage biomass and leaf area in wheat. The use of leaf area for estimation of conventional physiological parameters such as leaf area index will be useful in crop improvement programmes. As unique research, it included many biologically relevant image-based traits, including NIR mean gray value, RGB colour class data and mean RGB-derived indices. This work showed the vast potential for future applicability in discovering novel QTLs for biomass and growth-related traits at different phenological stages. Precise phenotyping of biomass at multiple time points in the plant life cycle will serve as a seed for the functional mapping of dynamic traits. It also has future implications in characterizing and quantifying nutrient deficiency effects at the different phenological time scale.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

The manuscript was reviewed and approved for publication by all authors. VC, SK and BS: Conceptualization and designing of experiment. BS: Final data analysis, interpretation and writing-

original draft. BS, SK, DV, SA, PS and GP: Experiment setup, data collection and data input. BS, AE and ND: Formal data analysis and visualization. VC, SK, HK, RS, DR, MD and LS: Performance evaluation and review of whole experiment. VC, SK, HK, RS, DR, MD and LS: Revision of manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by National Agricultural Science Fund, ICAR, New Delhi, Grant Nos. NFBSFARA/Phen-2015, NASF/Phen-6005/2016–17, and part of this research was supported by the grant from Bill and Melinda Gates Foundation (OPP1194767).

Acknowledgments

All authors acknowledge ICAR-Indian Agricultural Research Institute, New Delhi, for providing research facilities to conduct the research. BS is thankful to the CSIR for Senior Research Fellowship during his Ph.D. studies.

References

- Adak, A., Conrad, C., Chen, Y., Wilde, S. C., Murray, S. C., Anderson, S. L.II, et al. (2021). Validation of functional polymorphisms affecting maize plant height by unoccupied aerial systems discovers novel temporal phenotypes. *G3* 11 (6), 1–13. doi: 10.1093/g3journal/jkab075
- Al-Tamimi, N., Brien, C., Oakey, H., Berger, B., Saade, S., Ho, Y. S., et al. (2016). Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. *Nat. Commun.* 7 (1), 13342. doi: 10.1038/ncomms13342
- Arruda, M. P., Brown, P. J., Lipka, A. E., Krill, A. M., Thurber, C., and Kolb, F. L. (2015). Genomic selection for predicting fusarium head blight resistance in a wheat breeding program. *J. Plant Genome Sci.* 8 (3), 1–12. doi: 10.3835/plantgenome2015.01.0003
- Arvidsson, S., Pérez-Rodríguez, P., and Mueller-Roeber, B. A. (2011). Growth phenotyping pipeline for arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytol.* 191, 895–907. doi: 10.1111/j.1469-8137.2011.03756.x
- Asif, M. A., Schilling, R. K., Tilbrook, J., Brien, C., Dowling, K., Rabie, H., et al. (2018). Mapping of novel salt tolerance QTL in an Excalibur × Kukri doubled haploid wheat population. *Theor. Appl. Genet.* 131 (10), 2179–2196. doi: 10.1007/s00122-018-3146-y
- Baranowski, P., Jedryczka, M., Mazurek, W., Babula-Skowronska, D., Siedliska, A., and Kaczmarek, J. (2015). Hyperspectral and thermal imaging of oilseed rape (*brassica napus*) response to fungal species of the genus *alternaria*. *PLoS One* 10 (3), e0122913. doi: 10.1371/journal.pone.0122913
- Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* 39, 79–87. doi: 10.1016/j.jag.2015.02.012
- Busemeyer, L., Ruckelshausen, A., Möller, K., Melchinger, A. E., Alheit, K. V., Maurer, H. P., et al. (2013). Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Sci. Rep.* 3, 2442. doi: 10.1038/srep02442
- Buxbaum, N., Lieth, J. H., and Earles, M. (2022). Non-destructive plant biomass monitoring with high spatio-temporal resolution via proximal RGB-d imagery and end-to-end deep learning. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.758818
- Camargo, A. V., Mackay, I., Mott, R., Han, J., Doonan, J. H., Askew, K., et al. (2018). Functional mapping of quantitative trait loci (qtls) associated with plant performance in a wheat magic mapping population. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00887
- Campbell, M. T., Knecht, A. C., Berger, B., Brien, C. J., Wang, D., and Walia, H. (2015). Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. *Plant Physiol.* 168 (4), 1476–1489. doi: 10.1104/pp.15.00450
- Chang, T. G., Chang, S., Song, Q. F., Perveen, S., and Zhu, X. G. (2019). Systems models, phenomics and genomics: three pillars for developing high-yielding photosynthetically efficient crops. *Isp* 1 (1), 1093. doi: 10.1093/insilicoplants/diy003
- Chao, H., Raboanahiry, N., Wang, X., Zhao, W., Chen, L., Guo, L., et al. (2019). Genetic dissection of harvest index and related traits through genome-wide quantitative trait locus mapping in *Brassica napus* L. *Breed. Sci.* 69 (1), 104–116. doi: 10.1270/jsbbs.18115
- Chen, D., Fu, L. Y., Hu, D., Klukas, C., Chen, M., and Kaufmann, K. (2018a). The HTPmod shiny application enables modeling and visualization of large-scale biological data. *Commun. Biol.* 1 (1), 89. doi: 10.1038/s42003-018-0091-x
- 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. *Plant Cell* 26 (12), 4636–4655. doi: 10.1105/tpc.114.129601
- Chen, D., Shi, R., Pape, J.-M., Neumann, K., Arend, D., Graner, A., et al. (2018b). Predicting plant biomass accumulation from image-derived parameters. *Gigascience* 7 (2), 1–3. doi: 10.1093/gigascience/giy001
- Chen, J., Zhu, M., Liu, R., Zhang, M., Lv, Y., Liu, Y., et al. (2020). BIOMASS YIELD 1 regulates sorghum biomass and grain yield via the shikimate pathway. *J. Exp. Bot.* 71 (18), 5506–5520. doi: 10.1093/jxb/eraa275
- Elangovan, A., Duc, N. T., Raju, D., Kumar, S., Singh, B., Vishwakarma, C., et al. (2023). Imaging Sensor-Based High-Throughput Measurement of Biomass Using Machine Learning Models in Rice. *Agriculture* 13(4), 852. doi: 10.3390/agriculture13040852
- Esposito, S., Carputo, D., Cardi, T., and Tripodi, P. (2019). Applications and trends of machine learning in genomics and phenomics for next-generation breeding. *Plants J.* 9 (1), 34. doi: 10.3390/plants9010034
- Giraldo, P., Benavente, E., Manzano-Agugliaro, F., and Gimenez, E. (2019). Worldwide research trends on wheat and barley: a bibliometric comparative analysis. *Agronomy* 9 (7), 352. doi: 10.3390/agronomy9070352
- Golzarian, M. R., Frick, R. A., Rajendran, K., Berger, B., Roy, S., Tester, M., et al. (2011). Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods* 7 (1), 2. doi: 10.1186/1746-4811-7-2
- Guo, Z., Yang, W., Chang, Y., Ma, X., Tu, H., Xiong, F., et al. (2018). Genome-wide association studies of image traits reveal genetic architecture of drought resistance in rice. *Mol. Plant* 11 (6), 789–805. doi: 10.1016/j.molp.2018.03.018
- Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., et al. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods* 15 (1), 1–19. doi: 10.1186/s13007-019-0394-z

Conflict 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.

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.2023.1214801/full#supplementary-material>

- Hensman, J., Fusi, N., and Lawrence, N. D. (2013). Gaussian Processes for big data. *arXiv preprint arXiv:1309.6835*. doi: 10.48550/arXiv.1309.6835
- Jiang, L., Sun, L., Ye, M., Wang, J., Wang, Y., Bogard, M., et al. (2019). Functional mapping of n deficiency-induced response in wheat yield-component traits by implementing high-throughput phenotyping. *Plant J.* 97 (6), 1105–1119. doi: 10.1111/tj.14186
- Jin, X., Shi, C., Yu, C. Y., Yamada, T., and Sacks, E. J. (2017). Determination of leaf water content by visible and near-infrared spectrometry and multivariate calibration in miscanthus. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00721
- Krauth, K., Bonilla, E. V., Cutajar, K., and Filippone, M. (2016). AutoGP: exploring the capabilities and limitations of Gaussian process models. *arXiv preprint arXiv:1610.05392*. doi: 10.48550/arXiv.1610.05392
- Koehrsen, W. (2018). Introduction to Bayesian Linear Regression - Towards Data Science. *Medium*.
- Kumar, S., Raju, D., Sahoo, R. N., and Chinnusamy, V. (2016). Phenomics: unlocking the hidden genetic variation for breaking the barriers in yield and stress tolerance. *Indian J. Plant Physiol.* 21, 409–419. doi: 10.1007/s40502-016-0261-0
- Li, Y., and Wu, R. (2010). Functional mapping of growth and development. *Biol Res Camb Philos Soc* 85(2), 207–216. doi: 10.1111/j.1469-185X.2009.00096.x
- Lyra, D. H., Virlet, N., Sadeghi-Tehran, P., Hassall, K. L., Wingen, L. U., Orford, S., et al. (2020). Functional QTL mapping and genomic prediction of canopy height in wheat measured using a robotic field phenotyping platform. *J. Exp. Bot.* 71 (6), 1885–1898. doi: 10.1093/jxb/erz545
- McLeay, R. C., Lesluyes, T., Cuellar Partida, G., and Bailey, T. L. (2012). Genome-wide in silico prediction of gene expression. *Bioinformatics* 28 (21), 2789–2796. doi: 10.1093/bioinformatics/bts529
- Misra, T., Arora, A., Marwaha, S., Chinnusamy, V., Rao, A. R., Jain, R., et al. (2020). SpikeSegNet-a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods* 16 (1), 40. doi: 10.1186/s13007-020-00582-9
- Neeraj, P., Dinesh Kumar, S., and Sundip, K. (2022). Breaking Yield Ceiling in Wheat: Progress and Future Prospects. In A. Mahmood-ur-Rahman (Ed.), *Wheat* (pp. Ch. 2). IntechOpen. doi: 10.5772/intechopen.102919
- Neilson, E. H., Edwards, A. M., Blomstedt, C. K., Berger, B., Möller, B. L., and Gleadow, R. M. (2015). Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. *J. Exp. Bot.* 66 (7), 1817–1832. doi: 10.1093/jxb/eru526
- 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 (10), 1980–1996. doi: 10.1111/pce.12516
- Ogutu, J. O., Schulz-Streeck, T., and Piepho, H. P. (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proc.* 6 (2), 1–6. doi: 10.1186/1753-6561-6-S2-S10
- Parent, B., Shahinnia, F., Maphosa, L., Berger, B., Rabie, H., Chalmers, K., et al. (2015). Combining field performance with controlled environment plant imaging to identify the genetic control of growth and transpiration underlying yield response to water-deficit stress in wheat. *J. Exp. Bot.* 66, 5481–5492. doi: 10.1093/jxb/erv320
- Pérez-Valencia, D. M., Rodríguez-Álvarez, M. X., Boer, M. P., Kronenberg, L., Hund, A., Cabrera-Bosquet, L., et al. (2022). A two-stage approach for the spatio-temporal analysis of high-throughput phenotyping data. *Sci. Rep.* 12 (1), 3177. doi: 10.1038/s41598-022-06935-9
- Porker, K., Straight, M., and Hunt, J. R. (2020). Evaluation of G × e × m interactions to increase harvest index and yield of early sown wheat. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00994
- Pottier, J., and Jabot, F. (2017). Non-destructive biomass estimation of herbaceous plant individuals: a transferable method between contrasted environments. *Ecol. Indic.* 72, 769–776. doi: 10.1016/j.ecolind.2016.08.030
- Qiao, K., Zhu, W., Xie, Z., and Li, P. (2019). Estimating the seasonal dynamics of the leaf area index using piecewise LAI-VI relationships based on phenophases. *Remote Sens.* 11, 689. doi: 10.3390/rs11060689
- Rabab, S., Breen, E., Gebremedhin, A., Shi, F., Badenhurst, P., Chen, Y.-P. P., et al. (2021). A new method for extracting individual plant bio-characteristics from high-resolution digital images. *Remote Sens.* 13 (6), 1212. doi: 10.3390/rs13061212
- Rahaman, M. M., Ahsan, M. A., Gillani, Z., and Chen, M. (2017). Digital biomass accumulation using high-throughput plant phenotype data analysis. *Integr. Bioinform.* 14 (3), 1–13. doi: 10.1515/ijib-2017-0028
- Ramadas, S., Kiran Kumar, T. M., and Pratap Singh, G. (2020). Wheat production in india: trends and prospects. *Recent Advances in Grain Crops Research*. Ed. F. Shah IntechOpen. doi: 10.5772/intechopen.86341
- Rauf, S., Zaharieva, M., Warburton, M. L., Zhang, P.-Z., Al-Sadi, A. M., Khalil, F., et al. (2015). Breaking wheat yield barriers requires integrated efforts in developing countries. *J. Integr. Agric.* 14 (8), 1447–1474. doi: 10.1016/S2095-3119(15)61035-8
- R Core Team (2021). *R: a language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: <https://www.R-project.org/>.
- Reynolds, M. P., Pask, A. J. D., Hoppitt, W. J. E., Sonder, K., Sukumaran, S., Molero, G., et al. (2017). Strategic crossing of biomass and harvest index–source and sink–achieves genetic gains in wheat. *Euphytica* 213 (11), 257. doi: 10.1007/s10681-017-2040-z
- Roth, L., Rodríguez-Álvarez, M. X., van Eeuwijk, F., Piepho, H.-P., and Hund, A. (2021). Phenomics data processing: a plot-level model for repeated measurements to extract the timing of key stages and quantities at defined time points. *Field Crops Res.* 274, 108314. doi: 10.1016/j.fcr.2021.108314
- Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21 (2), 110–124. doi: 10.1016/j.tplants.2015.10.015
- Song, L., Huang, S. C., Wise, A., Castanon, R., Nery, J. R., Chen, H., et al. (2016). A transcription factor hierarchy defines an environmental stress response network. *Science* 354 (6312), 97–116. doi: 10.1126/science.aag1550
- 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 (3), 633–645. doi: 10.1016/j.cj.2021.03.015
- Strock, C. F., Schneider, H. M., and Lynch, J. P. (2022). High-throughput phenotyping of plant anatomy. *Trends Plant Sci.* 27 (6), 520–523. doi: 10.1016/j.tplants.2022.02.009
- Tackenberg, O. (2007). A new method for non-destructive measurement of biomass, growth rates, vertical biomass distribution and dry matter content based on digital image analysis. *Ann. Bot.* 99 (4), 777–783. doi: 10.1093/aob/mcm009
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., and Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Curr. Biol.* 27 (15), 770–783. doi: 10.1016/j.cub.2017.05.055
- Uddin, S., Haque, I., Lu, H., Moni, M. A., and Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Sci. Rep.* 12, 6256. doi: 10.1038/s41598-022-10358-x
- Van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., et al. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci.* 282, 23–39. doi: 10.1016/j.plantsci.2018.06.018
- Vasquez, M. M., Hu, C., Roe, D. J., Chen, Z., Halonen, M., and Guerra, S. (2016). Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. *BMC Med. Res. Methodol.* 16, 1–19. doi: 10.1186/s12874-016-0254-8
- Wu, S., Deng, L., Guo, L., and Wu, Y. (2022). Wheat leaf area index prediction using data fusion based on high-resolution unmanned aerial vehicle imagery. *Plant Methods* 18 (1), 1–16. doi: 10.1186/s13007-022-00899-7
- Wu, R., and Lin, M. (2006). Functional mapping - how to map and study the genetic architecture of dynamic complex traits. *Nat. Rev. Genet.* 7 (3), 229–237. doi: 10.1038/nrg1804
- Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *Theor. Appl. Genet.* 129 (4), 653–673. doi: 10.1007/s00122-016-2691-5
- Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. B., et al. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13 (2), 187–214. doi: 10.1016/j.molp.2020.01.008
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5 (1), 5087. doi: 10.1038/ncomms6087