Unraveling relationship between physiological and seed-yield traits in wheat (Triticum aestivum L.) through genotype–trait interactions

Ali Omrani

a.omrani@areeo.ac.ir

Agricultural Research & Education Organization

Seyed Habib Shojaei

Agricultural Research & Education Organization

Hossein Abbasi Holasou

Agricultural Research & Education Organization

Saeed Omrani

Sejong University

Mahvash Afshari

Sejong University

Mohammad Hosein Bijeh Keshavarzi

Shahed University

Fatih Demirel

Iğdır Üniversitesi

Jan Bocianowski

Poznań University of Life Sciences

Aras Türkoğlu

Necmettin Erbakan University

Research Article

Keywords: Wheat, Genotype × Environment interaction, Physiological traits, Machine learning, Yield prediction

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10401690/v1

Additional Declarations: No competing interests reported.

Unraveling relationship between physiological and seed-yield traits in wheat (*Triticum aestivum* L.) through genotype–trait interactions

Ali Omrani[*1], Seyed Habib Shojaei[2], Hossein Abbasi Holasou[3], Saeed Omrani[4], Mahvash Afshari[4], Mohammad Hosein Bijeh Keshavarzi[5], Fatih Demirel[6], Jan Bocianowski[7], Aras Türkoğlu[8]

[1] Crop and Horticultural Science Research Department, Ardabil Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Moghan 5695157451, Iran (a.omrani@areeo.ac.ir) (A.O.)
[2] Crop and Horticultural Science Research Department, Markazi Agricultural and Natural Resources Research and Education Center, Agricultural Research Education and Extension Organization (AREEO), Arak, Iran (h.shojaei@areeo.ac.ir) (S.H.S)
[3] Crop and Horticultural Science Research Department, Kurdistan Agricultural and Natural Resources Research and Education Center, Agricultural Research Education and Extension Organization (AREEO), Sanandaj, Iran. (h.abbasi@areeo.ac.ir) (H.A.H)
[4] Department of Bio-resource Engineering, College of Life Sciences, Sejong University, 209 Neungdong-ro, Gwanjing-gu, Seoul 05006, Republic of Korea (s.omrani70@gmail.com) (S.O.) (M.A.)
[5] Department of Agronomy, College of Agricultural Sciences, Shahed University, Tehran, Iran (mohammadhosein.keshavarzi@shahed.ac.ir) (M.H.B.K.)
[6] Department of Agricultural Biotechnology, Faculty of Agriculture, Igdır University, 76000 Igdir, Türkiye (fatih.demirel@igdir.edu.tr) (F.D.)
[7] Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland; jan.bocianowski@up.poznan.pl (J.B.)
[8] Department of Field Crops, Faculty of Agriculture, Necmettin Erbakan University, Konya 42310, Türkiye; aras.turkoglu@erbakan.edu.tr (A.T.)
***** Correspondence: a.omrani@areeo.ac.ir (A.O.)

Abstract

Wheat (*Triticum aestivum* L.) is one of the world's most important cereal crops, making research aimed at enhancing yield and improving agronomic and physiological traits a key priority. This study examined genotype × environment (G×E) interactions across five distinct regions. The results revealed significant G×E effects for most traits, except for variable fluorescence, chlorophyll index, and chlorophyll *a*, which exhibited minimal interaction. Mean comparisons of G×E interactions for grain yield identified G6 and G2 in the Damavand region and G9 in Qazvin as the highest-performing genotypes. Polygon plot analysis across all regions indicated that G21, G14, G10, and G5 were the most stable and desirable genotypes, while genotype ranking based on ideal performance favored G2, G12, and G10. Eigenvalue decomposition indicated that the first ten principal components accounted for more than 84% of the total variance, underscoring strong patterns in trait variation. Collectively, analyses across the studied regions identified G2, G10, and G12 as the most promising genotypes for future breeding programs. In predictive modeling, the extreme gradient boosting (XGBoost) algorithm demonstrated superior performance compared to random forest (RF) and support vector machine (SVM) models. For the test dataset, XGBoost achieved R^2 values of 0.652 (Karaj), 0.628 (Qazvin), and 0.698 (Damavand), along with RMSE values of 0.084, 0.067, and 0.114, respectively. The mean absolute percentage error (MAPE) remained consistently low (3.014%, 2.503%, and 2.046%), confirming the robustness of the model. In conclusion, this study identified high-performing wheat genotypes (G21, G14, G10, and G5) with stable adaptability across diverse environments, making them strong candidates for breeding programs. Furthermore, the successful implementation of machine learning (RF, SVM, and XGBoost) highlights their potential to enhance yield prediction and optimize genotype selection strategies in wheat improvement programs.

Keywords: Wheat, Genotype × Environment interaction, Physiological traits, Machine learning, Yield prediction.

1. Introduction

Wheat (*Triticum aestivum* L.) is one of the world's most strategic crops, providing a substantial proportion of dietary protein and energy to the global population (Arzani and Ashraf 2017). Its high nutritional value, affordability, widespread availability, and compatibility with the human digestive system have established its central role in the diets of approximately 4.6 billion people worldwide (Abbasi Holasou et al. 2023). Globally, wheat is cultivated at approximately 217 million hectares, yielding around 765 million tons of grain annually. However, driven by rapid population growth, global demand for wheat continues to increase. In Iran, wheat is cultivated on approximately 6.41 million hectares, of which 39% is irrigated and 61% is rainfed. According to the Food and Agriculture Organization of the United Nations (FAO), a global shortfall of approximately 100 million tons of wheat was recorded in 2020. Projections by the same organization indicate that by 2050, annual global wheat demand will reach approximately 840 million tons. Given the limited potential for expanding arable land, enhancing yield per unit area represents the most viable strategy to meet rising global demand and ensure food security. However, wheat production faces significant challenges due to the intensifying impacts of global climate change in recent years.

One of the most effective strategies to enhance wheat production is increasing yield per unit area. Evaluating the genetic diversity among wheat genotypes based on physiological traits associated with yield and its components enables breeders to identify the true genetic potential and performance capacity of various genotypes. Understanding the variation among wheat genotypes in key physiological traits, as well as the relationships among these traits, is essential for selecting suitable parent lines in breeding programs aimed at yield improvement. Such insights are critical for the development and advancement of high-yielding wheat cultivars.

Leaf area is one of the key determinants of a plant's photosynthetic capacity, as photosynthetic potential and growth rate are strongly correlated with leaf surface area (Navabpour et al. 2020). The total dry matter produced by a crop reflects the efficiency of the plant canopy in capturing and utilizing solar radiation throughout the growing season. To achieve maximum light interception and optimal photosynthesis efficiency, the crop canopy must maintain sufficient leaf area to ensure uniform and complete ground coverage. The crop growth rate (CGR) quantifies the amount of dry matter accumulated per unit land area per unit time, representing the combined effects of photosynthesis and respiration. The relative growth rate (RGR) indicates the rate at which a plant accumulates dry matter relative to its existing biomass over time. In most agricultural crops, RGR tends to decline as the plant progresses through its life cycle. Positive and significant correlations have been reported between CGR and RGR at the flowering stage and seed yield, while the net assimilation rate (NAR) has shown a positive correlation with RGR (Karimi and Siddique 1991). Moreover, a strong relationship has been observed between total dry matter production (TDM) and relative water content (Ashraf and Mehmood 1990). Chlorophyll fluorescence measurements are widely used to assess the photosynthetic performance of plants under field conditions (Kocheva et al. 2004). This technique is grounded in physiological principles that evaluate the efficiency of light harvesting and energy absorption, specifically assessing the functional status of photosystem II in chlorophyll (Maxwell and Johnson 2000). SPAD and Fv/Fm measurements are rapid, non-invasive methods that have proven effective as indices for assessing plants physiological status (Batool 2019). Leaf relative water content (RWC) is also considered a crucial indicator of plant water status, reflecting the equilibrium between water supply to leaf tissues and transpiration rate (Lugojan and Ciulca 2011). Research investigation the effects of water deficit stress on seedling biomass and physio-chemical traits across different wheat species demonstrated that the Fv/Fm parameter serves as one of the most reliable criteria for screening tolerant genotypes at the seedling stage. Collectively, these parameters facilitate rapid and early phenotyping of both domesticated and wild accessions, thereby enabling the identification of naturally occurring genetic variation (Pour-Aboughadareh et al. 2019).

The GGE biplot method was originally developed to analyzing data from multi-environment experiments (Yan and Kang 2003) but is equally applicable to any two-way data structure, such as genotype × trait interaction (Yan and Rajcan 2002). The GGE biplot technique enables the simultaneous graphical evaluation of genotype performance and genotype–environment interactions (Yan and Kang 2003). Genotypic assessment should be based on multiple traits align with breeding objectives. Genotype × trait (GT) data can be effectively visualized using a GT biplot, which is analogous to the GGE biplot except that the data must be standardized to eliminate the influence of differing measurement units among traits. Scaling by the standard deviation (SD) of each trait is the most common method in multi-trait data analysis. For repeated measurements, a scaling method incorporating both SD and weighting by the square

root of heritability (*h*) is preferred. For example, Qaseem et al. applied multivariate methods to identify three superior wheat genotypes out of 30, exhibiting higher yields and yield component values suitable for cultivation in specific regions of Pakistan (Qaseem et al. 2021). Similarly, Omrani et al. (2022) applied multivariate methods and graphical approaches to identify superior genotypes for multiple traits across diverse environments. Shojaei et al. employed graphical analysis to investigate genotype × trait interactions in maize hybrids, successfully identifying optimal hybrids with desirable trait combinations (Shojaei et al. 2022b). While discoveries in agricultural research are significant, robust statistical analyses that yield predictive insights are equally critical for translating findings into practical applications (Elfanah et al. 2023). Machine learning (ML) enables the development of mathematical models capable of generating predictions based on datasets comprising independent (inputs) and dependent (outputs) variables (Demirel et al. 2023a). ML techniques address the challenges of modeling complex interactions and nonlinear relationships inherent in agricultural datasets (Zhou et al. 2017). Consequently, ML can overcome the limitations of conventional univariate analyses, providing a more comprehensive and precise interpretation of complex data structures (Aasim et al. 2022). Machine learning has demonstrated significant potential in various domains of agricultural science, including *in vitro* studies, germination research, and yield optimization, where it has been applied for parameter prediction and optimization (Çelik et al. 2018; Kirtis et al. 2022). Moreover, numerous studies have compared different forecasting models to determine the most effective prediction approaches. As a result, machine learning algorithms can process data rapidly and accurately, optimize resource utilization, handle diverse data types and experimental conditions, and generate innovative solutions characterized by speed, precision, efficiency, adaptability, and creativity (Türkoğlu et al. 2023).

This study aims to identify the most promising wheat genotypes based on grain yield and key physiological traits associated with yield performance. Additionally, it seeks to investigate the underlying relationships between yield and physiological traits using multivariate analytical methods. Furthermore, this research aims to develop and compare multiple predictive models employing machine learning algorithms to determine the most accurate approach for yield prediction based on the observed trait data.

2. Materials and Methods

2.1. Testing and data collection

To examine the relationships between grain yield and physiological traits, as well as the genotype × trait interaction, an experiment was conducted using a randomized complete block design (RCBD) with three replications. The study evaluated 23 wheat genotypes across three regions: Karaj, Damavand, and Qazvin. All planting, cultivation, and harvesting operations were performed following standard agronomic practices, and to minimize edge effects, sampling was carried out from the central rows of each plot. Table 1 presents the codes and names of the evaluated genotypes, while Table 2 provides the geographical coordinates and rainfall data for each study site.

The experimental plots consisted of four rows, each 2 meters in length and spaced 50 cm apart. Seeds were sown at 7 cm intervals within each row. Standard agronomic practices for land preparation, planting, crop management, and harvesting were rigorously followed. No additional agricultural inputs were applied to ensure uniform across all plots. Most physiological traits were measured prior to harvest, whereas grain yield was determined post-harvest. To minimize edge effects, sampling and data collection were conducted from the two central rows of each plot. For each trait, data were recorded from five randomly selected plants within these rows, and mean values were calculated from these observations. The evaluated traits included proline content (Pro, µg g^{-1} FW), chlorophyll *a* (COLa, mg g^{-1} FW), chlorophyll *b* (COLb, mg g^{-1} FW), total chlorophyll (COLto, mg g^{-1} FW), carotenoid (CAR, mg g^{-1} FW), growth rate (CGR, gm^{-2} day^{-1}), dry matter production (TDM, g $plant^{-1}$), osmotic potential (OP, Mpa), cell membrane stability (CMS, %), leaf temperature (LT, °C), leaf stomatal conductance (SC, mol m^{-2} s^{-1}), florescence of chlorophyll (FC), fluorescence min (F0), fluorescence max (FM), variable fluorescence (FV), the ratio of variable fluorescence to maximum fluorescence (maximum quantum efficiency of photosystem II) (FV/FM), chlorophyll index (CI), leaf water potential (LWP, Mpa), relative growth rate (RGR, g g^{-1} day^{-1}), relative water content (RWC, %), grain yield (GYD, kg ha^{-1}), net photosynthesis rate (NAR, g m^{-2} day^{-1}) and leaf surface index (LAI).

Proline content was determined following the method described by (Bates et al. 1973). For this, one gram of leaf tissue was homogenized in 10 ml of 3% sulphosalicylic acid and centrifuged at 3000 rpm at 4°C for 10 minutes. Two milliliters of the resulting supernatant were mixed with two milliliters of ninhydrin reagent (prepared by

dissolving 1.25 g of ninhydrin powder into 30 ml of glacial acetic acid) and two milliliters of 6 M phosphoric acid. Subsequently, two additional milliliters of pure glacial acetic acid were added to the mixture. The tubes were then incubated in a hot water bath for one hour. After incubation, four milliliters of toluene were added to each tube, which was then vortexed vigorously for 15–20 seconds. The upper colored phase was carefully separated, and its absorbance was measured at 520 nm using a UV–visible spectrophotometer (model UV 2100, Unico, USA). Proline concentration was calculated in micrograms per gram of fresh leaf tissue using a standard curve. Chlorophyll and carotenoid contents were measured according to the method of (Arnon 1949). Briefly, 0.1 g of leaf tissue was thoroughly ground in a mortar with three milliliters of 80% acetone, and the final volume of the extract was adjusted to 15 ml. The extract was then centrifuged at 5000 rpm for 10 minutes. Absorbance of the supernatant was measured at wavelengths of 480, 510, 645, and 663 nm using a spectrophotometer (Shimadzu UV-160), with 80% acetone used as the blank. Chlorophyll *a*, chlorophyll *b*, total chlorophyll, and carotenoid contents were calculated using the following equations:

$$\text{Chlorophyll } a\ (mg\ g^{-1}) = [12.7(A633) - 2.69(A645)] \cdot V/1000 \cdot W$$
$$\text{Chlorophyll } b\ (mg\ g^{-1}) = [22.9(A645) - 4.68(A633)] \cdot V/1000 \cdot W$$
$$\text{Total chlorophyll } (mg\ g^{-1}) = [20.2(A645) + 8.02(A633)] \cdot V/1000 \cdot W$$
$$\text{Carotenoid } (mg\ g^{-1}) = [A480 + 0.114(A633) - 0.638(A645)] \cdot V/1000 \cdot W$$

Chlorophyll index measurements were conducted using a handheld chlorophyll meter (SPAD-502, Minolta), with the sampling randomly initiated from the uppermost leaves of the plants.

The crop growth rate (CGR) was calculated according to the following formula, where *w*1 and *w*2 correspond to the plant dry weights at times *t*1 and *t*2, respectively and GA represents the ground area occupied by the plant. The unit of CGR is grams per square meter of land per day.

$$CGR = \frac{w2 - w1}{t2 - t1} \cdot \frac{1}{GA}$$

To measure chlorophyll fluorescence, fully expanded leaves located at the top of the plant canopy were selected. These leaves were dark-adapted for 30 minutes using specialized Handy-PEA clips. During the dark adaptation period, the reaction centers of the photosynthetic apparatus were fully opened. Subsequently, a saturating light pulse with a wavelength of 650 nm and an intensity of 3000 μmol photons m^{-2} s^{-1} was applied for four seconds, and various fluorescence parameters were recorded using the device's dedicated software (PEA Plus) (Moharramnejad and Shiri 2020). To determine the maximum photochemical efficiency of photosystem II (PSII), the following formula was applied:

$$\text{Maximum photochemical efficiency of photosystem 2} = \frac{Fv}{Fm},$$

where *Fv* is variable fluorescence and *Fm* is maximum fluorescence.

Leaf temperature was measured using an infrared thermometer on randomly selected plants. Leaf area was determined with a leaf area meter (Hitachi, Japan), and the leaf area index (LAI) was subsequently calculated as the ratio of total leaf area to the ground area occupied by the respective plants.

The relative water content (RWC) of leaf tissue was calculated using the following formula, according to the method of (Egert and Tevini 2002):

$$RWC\% = \frac{Fw - Dw}{Tw - Dw} \cdot 100,$$

where: *Fw* – fresh weight of the leaf, *Dw* – dry weight of the leaf, *Tw* – turgid weight (saturated weight) of the leaf.

2.2. Machine learning (ML) assessment

The objective of this study was to predict grain yield measured in the Karaj, Qazvin, and Damavand regions by developing predictive models based on observed agronomic parameters collected from the same regions. It is important to note that the machine learning (ML) models were constructed independently for each environment (Karaj, Qazvin, and Damavand). As the same set of 23 genotypes was evaluated simultaneously under uniform field conditions at each location, the environmental effect (E) and genotype × environment interaction (GEI) were not confounded in the modeling process. Consequently, each model was trained and tested using data from a single environment, allowing the algorithms to focus exclusively on genotypic and physiological variability associated with grain yield. This design

ensured that the resulting predictions reflected genuine genotypic performance within each environment rather than differences among sites. Three machine learning (ML) algorithms were employed: the Random Forest (RF) model (Breiman 2001), the support vector machines (SVM) model (Noble 2006), and the extreme gradient boosting (XGBoost) model (Chen and Guestrin, 2016). The performance of these models was evaluated using four primary metrics: root mean square error (RMSE), coefficient of determination (R^2), mean absolute percentage error (MAPE), and mean absolute deviation (MAD). The coefficient of determination (R^2) quantifies the proportion of variance in the observed data explained by the model (Equation 1). The mean squared error (MSE) assesses the average squared difference between predicted and actual values (Equation 2). The mean absolute percentage error (MAPE) represents the average of absolute percentage differences between predicted and observed values (Equation 3). Lastly, the mean absolute deviation (MAD) evaluates the average magnitude of prediction errors (Equation 4)(Eren et al. 2023). The dataset was randomly partitioned into training (70%) and testing (30%) subsets using the 'caret' package in R (Darwish et al. 2023). Given the relatively small number of genotypes (n = 23), Leave-One-Out Cross-Validation (LOOCV) was employed to ensure robust model evaluation and minimize the risk of overfitting (Türkoğlu et al. 2023b). In the LOOCV procedure, the model is iteratively trained on all genotypes except one, which is used for testing; this process is repeated until each genotype has served once as the validation sample (Wong 2015). This approach provides a reliable estimation of model generalization, particularly for small datasets. The LOOCV method was implemented using the *train Control (method = "LOOCV")* function from the *caret* package (R version 4.3.2).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_{ip}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_{ip}\right)^2}1 \quad (2)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_{ip}}{y_i}\right| \cdot 100 \quad (3)$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_{ip}\right| \quad (4)$$

The variable n denotes the number of observations within the training or testing subset of the dataset. The variable y_i represents the observed (actual) value, y_{ip} indicates the predicted value, and $\bar{y}$ denotes the mean of the observed values. All machine learning model implementations and performance evaluations were conducted using R software (version 4.3.2) (Demirel et al. 2023a).

2.3. Statistical analysis of the experiment

To analyze the data collected from each region, a suite of statistical methods was applied, including combined analysis of variance (ANOVA), evaluation of genotype × environment (G×E) interaction effects on grain yield traits, polygon ("which-won-where") plots depicting genotypes performance across environments, and rank biplot analysis for classifying genotypes relative to the ideal genotype under different environmental conditions. Additionally, correlation diagrams were utilized to illustrate the relationships among the evaluated traits across experimental environments, while eigenvalue plots were used to visualize the decomposition of variance into principal components. Genotype distribution was further examined based on the first (PC1) and second (PC2) principal components across the different experimental locations.

Genotype × trait interactions were analyzed using the approach developed by Yan and Rajcan (2002), as outlined in Equation 5 below:

$$\frac{\alpha_{ij} - \beta_j}{\sigma_j} = \sum_{n=1}^{2}\lambda_n \xi_{in}\eta_{jn} + \varepsilon_{ij} = \sum_{n=1}^{2}\xi_{in}^{*}\eta_{jn}^{*} + \varepsilon_{ij} \quad (5)$$

where, α_{ij}: average amount of genotype i for every trait j, β_j: average amount of all the genotypes for the traits, σ_j: standard deviation of the trait j in the average genotypes, ε_{ij}: amount of genotype i remained in the trait j, λ_n: certain amount for the main element (PC_n), ξ_i: amount of PC_n for the genotype i, η_{jn}: amount of PC_n for the genotype j. Various researchers have used this equation in evaluating their genotypes and treatments (Kabiri et al. 2024; Mazloom et al. 2025; Pour-Aboughadareh et al. 2025; Shojaei et al. 2024).

3. Results and Discussion

3.1. Analysis of variance

Based on the combined analysis of variance (ANOVA) performed on the experimental data, the genotype effect was statistically significant ($p < 0.05$) for all traits except FC, COLa, and CAR, indicating substantial genetic diversity among the evaluated genotypes. This results suggests that genotypes possessing desirable characteristics can be effectively and selected for incorporation into regional breeding programs. The environmental effect was also significant for all traits except OP and RWC. Moreover, the genotype × environment (G×E) interaction was significant for all traits except FV, CI, and COLa. The highest and lowest coefficients of variation (CV) were observed for RGR (28.1%) and CGR (2.5%), respectively (Table 3). The relatively low CV values observed for most traits indicate a high degree of experimental precision across the study sites. The ANOVA results further demonstrated highly significant differences ($p < 0.05$) in the mean squares for genotypes, environments, and G×E interactions across the majority of evaluated traits. These outcomes indicate that wheat genotypes performance is strongly influenced by the cultivation environment and inter-annual variations in temperature and precipitation. Such results underscore the crucial role of environmental factors in the phenotypic expression of wheat genotypes and emphasize the necessity of conducting multi-year and multi-location trials. The significant G×E interaction suggests that genotype performance is not consistent across environments, highlighting the need to evaluate yield stability under diverse growing conditions. These results are consistent with prior studies reporting substantial G×E interactions for yield and yield-related traits in wheat genotypes (Akçura et al. 2005; Msundi et al. 2021), physiological parameters in bread wheat (Elfanah et al. 2023) and chlorophyll fluorescence (Fv/Fm) in quinoa (Hinojosa et al. 2019). Therefore, plant breeders should carefully consider climatic variability when selecting genotypes with stable performance to prevent the premature elimination of promising breeding lines or the unintended selection of inferior genetic material.

Based on the comparison analysis of the average genotype × environment interaction effect for grain yield, genotypes E2G6, E3G7 and E3G9 were identified as favorable, exhibiting high performance scores, whereas genotypes E3G10, E1G3, and E1G21 were classified as unfavorable due to their low performance ratings (Figure 1). Similarly, Shojaei et al. (2022a) and Omrani et al. (2022 and 2025) applied mean interaction effect comparisons to identify genotypes with superior adaptability and performance across environments.

3.2. Graphical analysis

A graphical analysis approach was employed to explore and interpret the relationships between grain yield and yield components in wheat genotypes. A polygon ("which-won-where") view was constructed to identify genotypes with superior performance for the evaluated traits (Figure 2). In this diagram, genotypes with the greatest distance from the origin were connected to form a polygon, enclosing the remaining genotypes within it. From the origin, perpendicular lines were drawn to each side of the polygon, dividing it into multiple sectors. The genotypes located at the vertices of each sector represent those with the highest performance for the traits associated with that sector (Yan and Kang 2003). Based on the polygon diagram for the Karaj location, genotypes G7, G19, G5, G15, G16, and G21 were identified as desirable due to their greater distance from the center of the biplot. Specifically, genotype G7 showed superior performance for COLto, COLa, LAI, CAR, and CGR traits; genotype G19 for RGR; genotypes G16 and G5 for RGR and CMS; genotype G15 for TDM and GYD; and genotype G21 for CI, FC, FM, and F0 (Figure 2a). In the Damavand location, the polygon view indicated that genotypes G18, G8, G9, G2, G10, G22, and G5 were the most favorable. Within this context, genotype G9 was superior for TDM, FC, and LWP; G10 and G8 for COLa; G22 for CGR and NAR; G5 for FV/FM; G2 for LAI and SC; and G18 for CI (Figure 2b). The polygon diagram for the Qazvin region similarly revealed that genotypes G14, G12, G4, G5, and G3 were the most promising. In detail, genotype G14 was superior for COLb, CAR, and CMS; G3 for CI and F0; G12 for COLa, GYD, RWC, RGR, COLto, and FV/FM; G4 for SC; and G5 for Pro, CGR, and FC (Figure 2c). The polygon view method has been widely used by numerous researchers for identifying high-performing genotypes in various crops, including maize (Mousavi et al. 2021; Shojaei et al. 2022a; Shojaei et al. 2020), wheat (Omrani et al. 2022 and 2025), sunflower (Ansarifard et al. 2020; Shojaei et al. 2022b).

According to the genotype ranking diagram, which is based on the concept of an ideal genotype, a reference line is drawn from the origin of the coordinate system through the mean point and extended in both directions. The optimal genotype is identified as the one located closest to the positive axis of this line while exhibiting the smallest

vertical deviation from it. In this graphical representation, the ideal genotype is denoted by the center of a set of concentric circles (indicated with an arrow), with all other genotypes ranked relative to this reference point. Based on the genotype ranking diagram for the Karaj region, genotypes G21, G2, G12, and G10 were identified as desirable, whereas genotypes G13, G16, G5, and G19 were considered less favorable. The genotypes are ordered from the most to the least desirable as follows G21> G2> G12> G10> G17> G8> G20> G14> G6> G3> G1> G15> G4> G11> G22> G7> G9> G18> G23> G16> G5> G19> G13 (Figure 3a).

Based on the ranking diagram for the Damavand region, genotypes G9 and G12 were identified as favorable, whereas genotypes G5, G15, G22, G2, and G17 were classified as less favorable. The genotypes are ordered from most to least desirable as follows G9> G12> G14> G3> G1> G11> G10> G6> G21> G23> G18> G7> G13> G16> G4> G20> G8> G10> G17> G2> G15> G22> G5 (Figure 3b). For the Qazvin region, the ranking diagram identified genotypes G10, G14, G3, G18, and G8 as favorable based on the ideal genotype concept, whereas genotypes G4, G12, G17, and G19 were classified as less favorable. The genotypes were ranked from the most to the least desirable as follows G10> G14> G3> G18> G8> G2> G9> G5> G20> G23> G22> G16> G21> G15> G13> G1> G7> G6> G11> G19> G17> G12> G4 (Figure 3c). Based on the ranking diagram for the Damavand region, genotypes G9 and G12 were identified as favorable, whereas genotypes G5, G15, G22, G2, and G17 were classified as less favorable. The genotypes are ordered from most to least desirable as follows G9> G12> G14> G3> G1> G11> G10> G6> G21> G23> G18> G7> G13> G16> G4> G20> G8> G10> G17> G2> G15> G22> G5 (Figure 3b). The ranking diagram for the Qazvin region identified genotypes G10, G14, G3, G18, and G8 as favorable based on the ideal genotype concept, while genotypes G4, G12, G17, and G19 were classified as less favorable. The genotypes are ranked from most to least desirable as follows G10> G14> G3> G18> G8> G2> G9> G5> G20> G23> G22> G16> G21> G15> G13> G1> G7> G6> G11> G19> G17> G12> G4 (Figure 3c).

The Damavand region, characterized by its high altitude, exhibits a cold climate with lower mean temperatures, a shorter growing season, and higher levels of photosynthetically active radiation (PAR). These environmental conditions can significantly influence physiological traits associated with photosynthetic efficiency. Genotypes G9 and G12, identified as superior performers in this region, exhibited not only high grain yield but also favorable values for Fv/Fm and RGR, indicating strong tolerance to cold stress and the physiological constraints imposed by low temperatures. In contrast, Karaj and Qazvin regions, which experience more moderate climates and longer growing seasons, offer more favorable conditions for the full expression of genotypic potential. In these regions, genotypes such as G21 in Karaj and G10 in Qazvin showed superior performance, particularly in crop growth rate (CGR). These differences in genotype rankings across the three environments are likely attributable to the climatic variations among the experimental locations.

By examining the genotype ranking diagrams based on the ideal genotype concept across all three evaluated regions, genotypes G2, G12, and G10 were identified as favorable, whereas genotypes G5, G19, and G17 were classified as unfavorable. These findings indicate that genotype selection was conducted not only on the basis of grain yield and yield components but also by incorporating related traits, including physiological characteristics, agronomic features, and spectral reflectance indices. These results are consistent with previous studies (Bakhshi and Shahmoradi 2023; Darwish et al. 2023; Elfanah et al. 2023; Mohammadi 2019; Sofi et al. 2021; Yan and Frégeau-Reid 2018). Therefore, it can be concluded that genotypes G2, G12, and G10 demonstrated superior performance according to multiple dimensions of biplot-based statistical analyses.

The correlations observed among traits reveal meaningful relationships that provide valuable insights for future studies, as visualized in the trait correlation biplot (Figure 4). In this graphical representation, the cosine of the angle between trait vectors indicates the strength and direction of their correlation: acute angles (<90°) signify positive correlations (+1), right angles (90°) indicate no correlation (0), and angles approaching 180° represent perfect negative correlations (–1), thus offering a quantitative framework for interpreting trait associations (Yan and Kang 2003). Based on the correlation biplot for the Karaj region, traits COLb, COLa, COLto, LAI, CAR, CGR, FV/FM, and SC were positively correlated with each other. Similarly, FV/FM, SC, FV, Pro, NAR, RWC, and RGR formed another positively correlated cluster. Traits GYD, CMS, TDM, and LT also showed positive interrelationships, while F0, FM, FC, CI, SPAD, and LWP grouped together with positive correlations (Figure 4a). The correlation diagram for the Damavand region likewise revealed positive correlations among FM, TDM, FC, COLb, OP, LWP, CAR, CMS, and

FV. Another cluster included FV, LT, COLto, COLa, GYD, and RGR. Traits RWC, FV/FM, CGR, NAR, and Pro were positively associated, as were OP, SC, LAI, CI, and F0 (Figure 4b). In the Qazvin region, positive correlations were observed among F0, CMS, CAR, COLb, LWP, FV, LAI, FM, and TDM. Another correlated group comprised FV/FM, COLto, RGR, RWC, GYD, and COLa, while SC, OP, and CI were positively correlated with CGR, FC, Pro, and LT (Figure 4c). Associations between yield and morpho-physiological traits have been broadly documented. For example, Gelalcha and Hanchinal (2013) and Msundi et al. (2021) reported significant positive correlations between yield, plant height, thousand kernel weight, the number of kernels per spike, and biomass in bread wheat. Sewore et al. (2023) further demonstrated that canopy temperature depression (CTD) traits measured at heading (CTDH), anthesis (CTDA), milking (CTDM), dough (CTDD), and ripening (CTDR) stages, as well as flag leaf chlorophyll content (SPAD502) at corresponding stages (SPADH, SPADA, SPADM, SPADD, SPADR), were positively and highly significantly correlated with one another and with other agronomic traits under both water regimes.

3.3. Principal components analysis (PCA)

According to the eigenvalue diagram for the Karaj environment, the first ten principal components accounted for more than 86% of the total data variance. Specifically, the first principal component (PC1) explained 14.4% of the variance, while the second principal component (PC2) accounted for over 12%, resulting in a combined contribution of approximately 26% of the total variance (Figures 5 and 6a). Based on the biplot illustrating the distribution of genotypes in relation to PC1 and PC2, genotypes G21, G2, G12, G17, G14, G3, G6, G20, G1, G15, G4, G16, G9, and G5 exhibited positive loadings on PC1. In addition, genotypes G7, G8, G22, G10, G17, G12, G2, G14, G3, G21, and G6 showed positive loadings on PC2. Notably, genotypes G21, G2, G12, G17, G14, G6, and G3 displayed positive associations with both the first and second principal components (Figure 6a).

According to the eigenvalue diagram for the Damavand environment, the first ten principal components accounted for over 84% of the total data variance. Among these, the first principal component (PC1) explained 12.54% of the variance, while the second principal component (PC2) accounted for slightly more than 12%. Collectively, PC1 and PC2 contributed 24.62% to the overall variance (Figures 5 and 6b). Based on the biplot of genotype distribution along the first two principal components, genotypes G7, G12, G9, G1, G21, G16, G6, G18, G23, G11, G19, and G10 exhibited positive loadings on PC1. Similarly, genotypes G4, G3, G13, G14, G20, G2, G7, G12, G9, and G1 showed positive loadings on PC2. Notably, genotypes G7, G12, G9, and G1 displayed positive contributions to both PC1 and PC2 (Figure 6b).

An analysis of the eigenvalue diagram based on the data from the Qazvin region revealed that the first ten principal components explained over 86% of the total data variance. Specifically, the first principal component (PC1) accounted for 16.5% of the variance, while the second principal component (PC2) explained 13.61%. Together, PC1 and PC2 contributed to more than 30% of the total variance (Figure 5). According to the biplot depicting genotype distribution along the first two principal components, genotypes G12, G11, G13, G15, G14, G9, G10, G8, and G3 showed positive loadings on PC1. In contrast, genotypes G4, G17, G19, G6, G7, G23, G21, G1, G15, G11, G13, and G12 exhibited positive loadings on PC2. Notably, genotypes G12, G11, G13, and G15 demonstrated positive contributions to both PC1 and PC2 (Figure 6c).

3.4. Machine learning analysis

Grain yield in wheat was predicted for the Karaj, Qazvin, and Damavand regions using three distinct machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). The results are presented in Table 4. Among the models evaluated, XGBoost demonstrated the lowest root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute deviation (MAD), indicating its superior predictive accuracy across all measured parameters in the training dataset. For the training data, the RMSE values for Karaj, Qazvin, and Damavand were 0.087, 0.053, and 0.094, respectively. Correspondingly, MAPE values were 3.127%, 2.650%, and 1.782%, while MAD values were 0.124, 0.789, and 0.024, respectively. The XGBoost model also achieved the highest coefficient of determination (R^2) for all three regions, with values of 0.648, 0.680, and 0.745 for Karaj, Qazvin, and Damavand, respectively. Model training was performed using the training dataset, and subsequent predictions were made on the test dataset to evaluate the generalization performance of each model. This approach facilitated assessment of the models' predictive capabilities on previously unseen data. Evaluation metrics

for the test dataset further confirmed the superiority of the XGBoost model. The R^2 values for Karaj, Qazvin, and Damavand were 0.652, 0.628, and 0.698, respectively. RMSE values were 0.084, 0.067, and 0.114; MAPE values were 3.014%, 2.503%, and 2.046%; and MAD values were 0.109, 0.740, and 0.028, respectively. Overall, the XGBoost model consistently outperformed the RF and SVM models based on all four performance metrics (R^2, RMSE, MAPE, and MAD). Notably, the highest predictive accuracy was observed in the Damavand region, where XGBoost achieved an R^2 of 0.698, indicating its strong potential for forecasting grain yield in that environment. These findings are consistent with prior research. For instance, Demirel et al. (2023a) demonstrated that machine learning techniques effectively capture nonlinear relationships between yield components and independent variables. Compared to conventional linear regression models, machine learning approaches offer significant improvements in predictive accuracy (Demirel et al. 2023a; Demirel et al. 2023b). According to Demirel et al. (2023a), machine learning enables comprehensive analysis of field data across multiple time points, thereby enhancing the optimization of crop development conditions and yield outcomes. Historical data, including previous agronomic practices and abiotic conditions, play a critical role in improving prediction models and informing future crop management strategies (Varga et al. 2023). Standard evaluation metrics such as RMSE, MAD, MAPE, and R^2 are widely used to identify the most effective and high-performing algorithms (Eren et al. 2023; Türkoğlu et al. 2023). In conclusion, the results of this study underscore the superior performance of the XGBoost model in predicting grain yield (GYL) across the Karaj, Qazvin, and Damavand regions. With R^2 scores ranging from 0.628 in Qazvin to 0.698 in Damavand, XGBoost demonstrated a high degree of predictive accuracy. These findings align with previous studies that have highlighted XGBoost's robustness in various agricultural applications, including the optimization of tissue culture parameters and prediction of regeneration outcomes (Demirel et al. 2023b; Şimşek et al. 2024; Türkoğlu et al. 2023). Developed by Chen and Guestrin (2016), the XGBoost algorithm based on gradient boosting decision trees is widely recognized for its high efficiency, scalability, and superior performance in both regression and classification tasks.

The superior performance of the XGBoost algorithm compared to RF and SVM can be attributed to its capacity to efficiently model nonlinear and complex genotype–trait relationships through sequential tree boosting. Unlike Random Forest, which generates multiple independent decision trees and averages their predictions, XGBoost constructs trees iteratively, with each subsequent tree learning from the residual errors of the its predecessors (Ilhan et al. 2024). This gradient boosting framework simultaneously reduces both bias and variance, resulting in enhanced accurate. Furthermore, XGBoost incorporates regularization techniques (L1 and L2 penalties) to mitigate overfitting and employs an internal mechanism for handling missing data, thereby improving its generalization performance across environments with variable inputs (Palaz et al. 2025). In contrast, the SVM performance may diminish when applied to high-dimensional or noisy agronomic datasets, particularly when kernel selection and parameter optimilization are challenging (Guido et al. 2024). These algorithmic advantages collectively explain why XGBoost consistently achieved higher R^2 values and lower RMSE, MAPE, and MAD across all studied environments.

Recent advances have underscored the transformative role of artificial intelligence (AI) and machine learning (ML) in plant breeding and the development of climate-resilient crops. Farooq et al. (2024) emphasized that the integration of AI-driven approaches into breeding pipelines can accelerate genetic gain by enhancing selection accuracy, capturing complex genotype × environment (G×E) interactions, and optimizing breeding decisions under variable climatic conditions. Similarly, Mushtaq et al. (2024) reported that ML-based predictive models such as XGBoost, Random Forest, and Support Vector Machines are increasingly applied to identify drought- and heat-tolerant wheat genotypes, integrating multi-omic, phenomic, and environmental datasets for more robust trait prediction . The present study aligns with these emerging frameworks by applying ML to evaluate genotypic performance across contrasting environments. Although the dataset was limited, the use of Leave-One-Out Cross-Validation (LOOCV) ensured reliable generalization, consistent with the small-sample modeling approaches discussed by Farooq et al. (2024). The superior performance of XGBoost observed in our results reflects its ability to iteratively reduce residual errors and capture non-linear trait interactions, supporting its growing relevance in predictive breeding workflows. Collectively, these findings demonstrate that combining conventional biometrical analyses (e.g., GGE, AMMI) with ML algorithms can strengthen the identification of stable and high-performing genotypes under climate variability, thereby contributing to the broader goal of data-driven and climate-smart wheat improvement.

The complete ranking and percentage contribution of yield-predictive variables identified by XGBoost for the Karaj, Qazvin, and Damavand environments are summarized in Supplementary Table S1. The variable importance analysis revealed that the most influential predictors of yield exhibited slight variation among environments; however, several physiological traits consistently contributed to yield formation. In Karaj, fluorescence minimum (F_0), relative water content (RWC), and leaf temperature exhibited the highest predictive importance, indicating that photosystem II efficiency and tissue water status were the main determinants of genotypic performance under this environment. These findings align with the multivariate analyses, where genotypes characterized by higher RWC and stable fluorescence parameters also clustered with high-yielding groups in the GGE and PCA biplots, suggesting that water retention capacity and chlorophyll fluorescence stability are reliable indicators of stress resilience.

In the Qazvin, the net photosynthesis rate, variable fluorescence (Fv), and osmotic potential were the top-ranked variables, emphasizing the role of photosynthetic efficiency and osmotic regulation under comparatively drier conditions. This pattern corresponded well with the multivariate separation of genotypes showing superior photosynthetic performance and osmotic adjustment, confirming that genotypes capable of maintaining positive carbon assimilation and turgor stability exhibited superior yield potential.

In the Damavand, the leaf surface index, dry matter production, and growth rate were the most influential predictors, indicating that canopy development and biomass accumulation were the dominant determinants of productivity in this cooler, high-elevation site. Similar patterns were evident in the multivariate analyses, where these growth-related traits had strong loadings on the first principal component, jointly contributing to the differentiation of high-yielding genotypes. Notably, relative water content and proline content also contributed moderately, indicating that both morphological and osmoprotective mechanisms jointly support yield stability in this environment.

Overall, the convergence between XGBoost-derived traits importance scores and the results of traditional multivariate analyses (PCA, GGE biplot, and correlation networks) validates both the robustness and the biological interpretability of the machine learning models. Traits associated with photosynthetic activity (F_0, Fv, photosynthesis rate), water status (RWC, osmotic potential, leaf temperature), and biomass formation (growth rate, dry matter) consistently emerged as key determinants across environments, confirming their physiological relevance to yield resilience in wheat. These results are in agreement with recent reports highlighting that the integration of ML approaches with classical biometrical analyses enhances the detection of genotype–trait–environment relationships and provides a more comprehensive understanding of yield stability mechanisms (Farooq et al. 2024; Mushtaq et al. 2024). In particular, the recurrent importance of traits related to photosynthetic efficiency and water-use dynamics supports earlier evidence that these physiological parameters serve as reliable proxies for stress tolerance and productivity under climate variability. For instance, Yang et al. (2024) demonstrated that XGBoost achieved superior accuracy in multi-sensor UAV-based yield prediction, highlighting its capacity to integrate complex phenotypic traits and environmental covariates. Similarly, Li et al. (2024) reported that integrating crop modeling with XGBoost improved dryland wheat yield prediction under variable irrigation regimes, illustrating the algorithm's strength in capturing nonlinear responses of physiological indicators to environmental stress . Furthermore, Yoosefzadeh-Najafabadi et al. (2025) emphasized that ML frameworks are becoming pivotal for genomic- and phenomic-assisted breeding by bridging classical multivariate statistics with data-driven approaches to reveal biologically meaningful trait hierarchies. The alignment between data-driven trait importance and well-established physiological mechanisms thus demonstrates that, when rigorously applied and validated, ML algorithms can complement traditional analytical frameworks, thereby strengthening decision-making processes in climate-resilient wheat breeding programs.

4. Conclusions

In this study, 23 agro-morphological traits of 23 bread wheat genotypes were evaluated across three distinct environments: Karaj, Damavand, and Qazvin (Iran). The analysis revealed significant differences ($p \leq 0.05$) among environments, genotypes, and their interactions for the majority of the evaluated traits. Based on the results of the polygon diagram analysis, genotypes G21, G14, G10, and G5 were identified as the most promising cultivars, exhibiting superior adaptability across all test locations. Nonetheless, further comprehensive multi-environmental trials are recommended prior to their commercial release. Moreover, the integration of machine learning (ML) techniques into optimization and predictive modeling processes enables more efficient and timely decision-making by capturing complex relationships between environmental conditions and grain yield in specific cultivars such as wheat. Machine

learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) have demonstrated strong predictive capabilities for key traits involved in complex biological systems. Further exploration of the interplay between ML models and target traits such as grain yield may uncover novel insights and strategies, ultimately enhancing the precision, efficacy, and sustainability of modern plant breeding programs.

Author Contributions

Conceptualization, A.O., S.H.S., H.A.H., S.O. and A.T.; methodology, A.O., H.A.H. and S.H.S.; software, A.O., H.A.H., S.O. and S.H.S.; validation, A.T., H.A.H., F.D. and A.O.; formal analysis, A.T.; investigation, H.A.H., A.T. and M.A.; resources, A. O., S.O., H.A.H. and S.H.S.; data curation, A.O., S.O., H.A.H. and S.H.S; writing—original draft preparation, A.T., A.O., H.A.H., S.O.; writing—review and editing, A.T., H.A.H., F.D., J.B. and A.O.; visualization, A.O., S.O., H.A.H., F.D., M.H.B.K. and S.H.S.; supervision, A.T.; project administration, A.O., H.A.H. and S.H.S; funding acquisition, A.O. and A.T. All authors have read and agreed to the published version of the manuscript.

Ethics declarations

Conflict of interest

The authors declare no conflict of interest.

Consent for publication

Informed consent was obtained from all individual participants included in the study.

Funding

This research received no external funding.

Data availability statement

Data is contained within the article.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aasim M, Katırcı R, Akgur O, Yildirim B, Mustafa Z, Nadeem MA, Baloch FS, Karakoy T, Yılmaz G (2022) Machine learning (ML) algorithms and artificial neural network for optimizing in vitro germination and growth indices of industrial hemp (*Cannabis sativa* L.). Industrial Crops and Products 181: 114801

Abbasi Holasou H, Alavi Kia SS, Mohammadi SA, Vahed MM (2023) Mixed linear models for the genetic inheritance of grain zinc and iron content, agronomic and biochemical traits in bread wheat under salinity stress. Biologia 78 (12): 3353-3365

Akçura M, Kaya Y, Taner S (2005) Genotype–environment interaction and phenotypic stability analysis for grain yield of durum wheat in Central Anatolian region. Turkish Journal of Agriculture and Forestry 29:369-375

Ansarifard I, Mostafavi K, Khosroshahli M, Reza Bihamta M, Ramshini H (2020) A study on genotype–environment interaction based on GGE biplot graphical method in sunflower genotypes (*Helianthus annuus* L.). Food Science & Nutrition 8 (7): 3327-3334

Arnon DI (1949) Copper enzymes in isolated chloroplasts. Polyphenoloxidase in beta vulgaris. Plant Physiol 24 (1): 1-15. https://doi.org/10.1104/pp.24.1.1

Arzani A, Ashraf M (2017) Cultivated ancient wheats (*Triticum* spp.): A potential source of health-beneficial food products. Compr Rev Food Sci Food Saf 16 (3): 477-488. https://doi.org/10.1111/1541-4337.12262

Ashraf M, Mehmood S (1990) Response of four Brassica species to drought stress. Environmental and Experimental Botany 30 (1): 93-100. https://doi.org/10.1016/0098-8472(90)90013-T

Bakhshi B, Shahmoradi SS (2023) Simultaneous selection of high-yielding and drought-tolerant barley landraces using GT, GYT and GYSI methodologies. Cereal Research Communications 51 (1): 237-248

Bates LS, Waldren R, Teare I (1973) Rapid determination of free proline for water-stress studies. Plant and soil 39 (1): 205-207

Batool A (2019) Assessing genetic variability and physiological traits conferring drought stress tolerance in cereals. University of Tasmania,

Breiman L (2001) Random forests. Machine learning 45 (1): 5-32

Çelik Ş, Boydak E, Firat R (2018) An analysis of factors affecting yield, oil production rate and plant height in sunflowers using selected data mining algorithms. Journal of Animal and Plant Sciences 28: 1085-1093

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. Paper presented at the proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA,

Darwish MA, Elkot AF, Elfanah AMS, Selim AI, Yassin MMM, Abomarzoka EA, El-Maghraby MA, Rebouh NY, Ali AM (2023) Evaluation of wheat genotypes under water regimes using hyperspectral reflectance and agro-physiological parameters via genotype by yield*trait approaches in Sakha Station, Delta, Egypt. Agriculture 13 (7). https://doi.org/10.3390/agriculture13071338

Demirel F, Eren B, Yilmaz A, Türkoğlu A, Haliloğlu K, Niedbała G, Bujak H, Jamshidi B, Pour-Aboughadareh A, Bocianowski J, Nowosad K (2023a) Prediction of grain yield in wheat by CHAID and MARS algorithms analyses. Agronomy 13 (6). https://doi.org/10.3390/agronomy13061438

Demirel F, Uğur R, Popescu GC, Demirel S, Popescu M (2023b) Usage of machine learning algorithms for establishing an effective protocol for the in vitro micropropagation ability of black chokeberry (*Aronia melanocarpa* (Michx.) Elliott). Horticulturae 9 (10). https://doi.org/10.3390/horticulturae9101112

Egert M, Tevini M (2002) Influence of drought on some physiological parameters symptomatic for oxidative stress in leaves of chives (*Allium schoenoprasum*). Environmental and Experimental Botany 48 (1): 43-49

Elfanah AMS, Darwish MA, Selim AI, Elmoselhy OMA, Ali AM, El-Maghraby MA, Abdelhamid MT (2023) Hyperspectral reflectance and agro-physiological traits for field identification of salt-tolerant wheat genotypes using the genotype by yield*trait biplot technique. Front Plant Sci 14: 1165113. https://doi.org/10.3389/fpls.2023.1165113

Eren B, Türkoğlu A, Haliloğlu K, Demirel F, Nowosad K, Özkan G, Niedbała G, Pour-Aboughadareh A, Bujak H, Bocianowski J (2023) Investigation of the influence of polyamines on mature embryo culture and DNA methylation of wheat (*Triticum aestivum* L.) using the machine learning algorithm method. Plants 12 (18). https://doi.org/10.3390/plants12183261

Farooq MA, Gao S, Hassan MA, Huang Z, Rasheed A, Hearne S, Prasanna B, Li X, Li H (2024) Artificial intelligence in plant breeding. Trends in Genetics 40 (10): 891-908

Gelalcha S, Hanchinal RR (2013) Correlation and path analysis in yield and yield components in spring bread wheat (*Triticum aestivum* L.) genotypes under irrigated condition in Southern India. African Journal of Agricultural Research 8 (24): 3186-3192. https://doi.org/10.5897/AJAR2013.6965

Guido R, Ferrisi S, Lofaro D, Conforti D (2024) An overview on the advancements of support vector machine models in healthcare applications: a review. Information 15 (4): 235

Hinojosa L, Kumar N, Gill KS, Murphy KM (2019) Spectral reflectance indices and physiological parameters in quinoa under contrasting irrigation regimes. Crop Science 59 (5): 1927-1944. https://doi.org/10.2135/cropsci2018.11.0711

Ilhan E, Koc AB, Kozat SS (2024) Exploiting residual errors in nonlinear online prediction. Machine Learning 113 (9): 6065-6091

Kabiri A, Zaefariyan F, Omrani A (2024) Genetic investigation of the resistance of promising wheat lines to virulence factors of stripe rust and leaf rust races. Cereal Research 13 (4): 351-366

Karimi MM, Siddique KHM (1991) Crop growth and relative growth rates of old and modern wheat cultivars. Australian Journal of Agricultural Research 42 (1): 13-20

Kirtis A, Aasim M, Katırcı R (2022) Application of artificial neural network and machine learning algorithms for modeling the in vitro regeneration of chickpea (*Cicer arietinum* L.). Plant Cell, Tissue and Organ Culture (PCTOC) 150 (1): 141-152

Kocheva K, Lambrev P, Georgiev G, Goltsev V, Karabaliev M (2004) Evaluation of chlorophyll fluorescence and membrane injury in the leaves of barley cultivars under osmotic stress. Bioelectrochemistry 63 (1-2): 121-124

Li Z, Nie Z, Li G (2024) Integrating crop modeling and machine learning for the improved prediction of dryland wheat yield. Agronomy 14 (4): 777

Lugojan C, Ciulca S (2011) Evaluation of relative water content in winter wheat. Journal of Horticulture, Forestry and Biotechnology 15 (2):173-177

Maxwell K, Johnson GN (2000) Chlorophyll fluorescence-a practical guide. Journal of Experimental Botany 51 (345): 659-668. https://doi.org/10.1093/jexbot/51.345.659

Mazloom P, Shojaie SH, Mousavi SMN (2025) Physiological and biochemical pathway: understanding their influence on maize grain yield through genotype–trait interactions. Scientific Reports 15 (1): 18329

Mohammadi R (2019) Genotype by yield*trait biplot for genotype evaluation and trait profiles in durum wheat. Cereal Research Communications 47 (3): 541-551. https://doi.org/10.1556/0806.47.2019.32

Moharramnejad S, Shiri M (2020) Study of genetic diversity in maize genotypes by ear yield and physiological traits. Journal of Crop Breeding 12: 30-40. https://doi.org/10.52547/jcb.12.35.30

Mousavi SM, Bojtor C, Illés Á, Nagy J (2021) Genotype by trait interaction (GT) in maize hybrids on complete fertilizer. Plants 10 (11). https://doi.org/10.3390/plants10112388

Msundi EA, Owuoche JO, Oyoo ME, Macharia G, Singh RP, Randhawa MS (2021) Identification of bread wheat genotypes with superior grain yield and agronomic traits through evaluation under rust epiphytotic conditions in Kenya. Sci Rep 11 (1): 21415. https://doi.org/10.1038/s41598-021-00785-7

Mushtaq MA, Ahmed HGM-D, Zeng Y (2024) Applications of artificial intelligence in wheat breeding for sustainable food security. Sustainability 16 (13): 5688

Navabpour S, Yamchi A, Bagherikia S, Kafi H (2020) Lead-induced oxidative stress and role of antioxidant defense in wheat (*Triticum aestivum* L.). Physiology and molecular biology of plants 26 (4): 793-802

Noble WS (2006) What is a support vector machine? Nature Biotechnology 24 (12): 1565-1567. https://doi.org/10.1038/nbt1206-1565

Omrani A, Omrani S, Khodarahmi M, Shojaei SH, Illés Á, Bojtor C, Mousavi SM, Nagy J (2022) Evaluation of grain yield stability in some selected wheat genotypes using AMMI and GGE biplot methods. Agronomy 12 (5). https://doi.org/10.3390/agronomy12051130

Omrani A, Omrani S, Shojaei SH, Abbasi Holasou H, Turkoglu A, Afzalifar A (2025) Analyzing wheat productivity: using GGE biplot and machine learning to understand agronomic traits and yield. Creal Research Communications, 53: 1861-1879. https://doi.org/10.1007/s42976-024-00615-2

Palaz EB, Demirel S, Popescu GC, Demirel F, Uğur R, Yaman M, Say A, Şimşek Ö, Tunç Y (2025) Refinement of surface sterilization protocol for in vitro olive (*Olea europaea* L.) shoot proliferation and optimizing by machine learning techniques. Horticulture, Environment, and Biotechnology 66(5): 1-16.

Pour-Aboughadareh A, Omidi M, Naghavi MR, Etminan A, Mehrabi AA, Poczai P, Bayat H (2019) Effect of water deficit stress on seedling biomass and physio-chemical characteristics in different species of wheat possessing the D genome. Agronomy 9 (9): 522

Pour-Aboughadareh A, Koohkan S, Omrani A, Marzooghian A, Gholipour A, Zali H, Kheirgoo M, Shahbazi-Homonloo K, Poczai P, Jamshidi B (2025) Combining multiple stability and adaptation models to analyze genotype-by-environment interactions for selection of stable barley genotypes with high yield performance. Agrosystems, Geosciences & Environment 8 (3): e70205

Qaseem MF, Qureshi R, Shaheen H, Waheed A (2021) Multivariate analysis for yield and proline content in wheat under lab and field conditions. Pak J Bot 53 (1): 227-239

Shojaei S, Mostafavi K, Bihamta M, Omrani A, Bojtor C, Illes A, Szabo A, Vad A, Nagy J, Harsányi E (2024) Selection of maize hybrids based on genotype× yield× trait (GYT) in different environments. Brazilian Journal of Biology 84: e272093

Shojaei S, Mostafavi K, Khosroshahli M, Bihamta M, Ramshini H (2022a) Evaluation of yield relationships and yield components in maize hybrids using multivariate and graphical methods in Karaj region. Journal of Crop Breeding 14: 174. https://doi.org/10.52547/jcb.14.41.174

Shojaei SH, Mostafavi K, Khosroshahli M, Reza Bihamta M, Ramshini H (2020) Assessment of genotype-trait interaction in maize (*Zea mays* L.) hybrids using GGT biplot analysis. Food Sci Nutr 8 (10): 5340-5351. https://doi.org/10.1002/fsn3.1826

Shojaei SH, Mostafavi K, Omrani A, Illés Á, Bojtor C, Omrani S, Mousavi SMN, Nagy J (2022b) Comparison of maize genotypes using drought-tolerance indices and graphical analysis under normal and humidity stress conditions. Plants (Basel) 11 (7): 942. https://doi.org/10.3390/plants11070942

Şimşek Ö, Dalda Şekerci A, Isak MA, Bulut F, İzgü T, Tütüncü M, Dönmez D (2024) Optimizing micropropagation and rooting protocols for diverse lavender genotypes: a synergistic approach integrating machine learning techniques. Horticulturae 10 (1): 52

Sofi P, Iram S, Ara A, Rehman K (2021) Comparative efficiency of GY*T approach over GT approach in genotypic selection in multiple trait evaluations: case study of common bean (*Phaseolus vulgaris*) grown under temperate himalayan conditions. Agricultural Research https://doi.org/10.1007/s40003-021-00577-5

Türkoğlu A, Bolouri P, Haliloğlu K, Eren B, Demirel F, Işık Mİ, Piekutowska M, Wojciechowski T, Niedbała G (2023) Modeling callus induction and regeneration in hypocotyl explant of fodder pea (*Pisum sativum* var. arvense L.) using machine learning algorithm method. Agronomy 13 (11): 2835. https://doi.org/10.3390/agronomy13112835

Varga I, Radočaj D, Jurišić M, Markulj Kulundžić A, Antunović M (2023) Prediction of sugar beet yield and quality parameters with varying nitrogen fertilization using ensemble decision trees and artificial neural networks. Computers and Electronics in Agriculture 212: 108076. https://doi.org/10.1016/j.compag.2023.108076

Yan W, Frégeau-Reid J (2018) Genotype by yield*trait (GYT) biplot: a novel approach for genotype selection based on multiple traits. Scientific Reports 8 (1): 8242. https://doi.org/10.1038/s41598-018-26688-8

Yan W, Kang M (2003) GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. https://doi.org/10.1201/9781420040371

Yan W, Rajcan I (2002) Biplot analysis of test sites and trait relations of soybean in ontario. Crop Sci 42 (1): 11-20. https://doi.org/10.2135/cropsci2002.1100

Yang S, Li L, Fei S, Yang M, Tao Z, Meng Y, Xiao Y (2024) Wheat yield prediction using machine learning method based on UAV remote sensing data. Drones 8 (7): 284

Yoosefzadeh-Najafabadi M, Xavier A, Eskandari M, Hesami M (2025) Machine learning after a decade: is it still a missing keystone in genomic-based plant breeding? Artificial Intelligence Review 58 (9): 260

Zhou L, Pan S, Wang J, Vasilakos AV (2017) Machine learning on big data: Opportunities and challenges. Neurocomputing 237: 350-361. https://doi.org/10.1016/j.neucom.2017.01.026

Table 1 Code and pedigree of wheat genotypes examined in the experiment.

Genotype Number	Pedigree
G1	*LAMAR* / / *BABAX* / 3 / *VORB* / 3 / *BORL*95
G2	*PF*70354 / / *CC*6010 / 4 / *KABY* / 4 / *BJY*
G3	*PUB*94 / 3 / *FRTL* / / *MUS* / 3* *BCN*
G4	*BATAVIA* / / *MTRWA*92.161 / 3* *PRINIA*
G5	*FRNCLN* * 2 / *KIRITATI* / / *TECUE* #1
G6	*Bilyana* / / *BAV*92 / 3 / *BRAMBLING*
G7	*ND*643 / / *ABC NAVO* / 3 / *C*80.1
G8	*T.TAU*.83 / / *EXCALIBURLG* / 3* *ANAPURNA*
G9	*ABC KLAUZIUS* / / *PBW*34 / 3 / *SQUARROSA*
G10	*RALITSA* / / *D*67.2 / *PARANAGEDIZ* / 3 / *GOO*
G11	*RIANNA* / / *PFAU* / *WEAVER* / / *KIRITATI*
G12	*VYARA* / / *KRONSTAD F*2004 / 3 / *KINGBI*
G13	*PFAU* / *ANETA* / / *JUCHI*
G14	*FAKTOY* / / *AGRI* / *NACIRA* / 3 / *SHARK* / *F*410*M*
G15	*GANSU* – 1 / *MEZGIT* – 4 / / *OGNYANA*
G16	*ABCALFIO* / *MEX*6 / / *BOW* / 3 / *GUN*91
G17	*REMESLINA* / / *TUR*3055 / 3 / *MOSKVICH*
G18	*APOGEI* / / *PRINIA* / 3 / *NAVG*54
G19	*QUARRO* / *ISKAD* / / *FRANCOLIN* #1* 2
G20	*PRYASPA* / / *HXL*75 / 2* *BAU*
G21	*A*47 / 415 / / *URES* / 5 / *VEE* / *LIRA* / / *BOW* / 3 / *BCN*
G22	*NEVEN* / / *FRET*2* 2 / 4 / *SNI* / *MUNIA*
G23	*TERVEL* / / *GK ARON* / *AG SECO* / 4 / 2* *MOLA*
G24	*ALISA* / / *MONARCA* / 2* *BAU*
G25	*Rakhshan* (*Control*)

Table 2 Geographical characteristics and rainfall of the areas evaluated in the experiment.

Area	Longitude	Latitude	Elevation MSL(m)	Average rainfall (mm)
Karaj	50° 57' 42.3000" E	35° 51' 21.3768" N	1312	295
Qazvin	50° 0' 11.5236" E	36° 16' 9.7068"N	1347	285
Damavand	52.109980"E	35.954836"N	2300	360

Table 3 Combined analysis of variance in terms of traits examined in the experiment

S.O. V	df	CGR	TDM	OP	Pro	CMS	LT	SC	FC	F0	FM	FV	FV/FM
Genotype	22	22.7*	30.2*	4.8*	7.9**	74.3*	0.22*	0.16*	3.4ns	1.05*	363.6*	48.7*	0.22*
Error 1	6	152.06	331.8	17.7	2.6	51.8	0.74	0.64	4.7	5.6	78.4	48.7	0.28
Environment	2	40.7*	124.8*	3.3ns	4.8*	34.6*	1.5**	1.2**	0.6*	6.2*	244.7*	15.8*	0.74*
Gen × Env	44	19.8*	43.8*	4.7*	5.6**	60.2*	0.25*	0.23*	3.7*	1.3*	286.5*	38.6ns	0.38*
Error 2	132	19.5	47.9	5.6	3.9	60.4	0.19	0.16	4.3	1.3	351.2	56.5	0.37
CV%		2.5	3.2	5.03	4.6	11.6	6.3	4.8	16.2	10.9	19.1	21.8	17.4

* , ** and ns: Significant (α= 5 %), highly significant (α= 1 %) and non-significant, respectively. CV: Coefficient of variation. Traits include growth rate (CGR), dry matter production (TDM), osmotic potential (OP), product proline (Pro), cell membrane stability (CMS), leaf temperature (LT), leaf stomatal conductance (SC), florescence of chlorophyll (FC), fluorescence min (F0), fluorescence max (FM), variable fluorescence (FV), the ratio of variable fluorescence to maximum fluorescence (maximum quantum efficiency of photosystem II) (FV/FM).

Table 3 (continued)

S.O.V	df	CI	LWP	RGR	RWC	COLa	COLb	COLto	CAR	GYD	NAR	LAI
Genotype	22	0.00006*	0.08*	34172*	75.7*	0.01^{ns}	0.002*	0.01*	0.002^{ns}	0.28*	0.72*	0.06*
Error 1	6	0.0007	0.03	42040	87.4	0.01	0.003	0.03	0.001	0.45	0.48	0.008
Environment	2	0.001*	0.03*	9365*	95.9^{ns}	0.008*	0.005*	0.009*	0.007*	0.84*	1.09*	0.01*
Gen × Env	44	0.0005^{ns}	0.06*	33422*	52.6*	0.006^{ns}	0.001*	0.009*	0.002*	0.3*	0.87*	0.03*
Error 2	132	0.0006	0.07	35029	81.8	0.009	0.002	0.009	0.002	0.3	0.96	0.03
CV%		24.1	17.05	28.1	10.8	19.1	29.7	14.8	29.1	23.4	11.8	15.2

* , ** and ns: Significant (α= 5 %), highly significant (α= 1 %) and non-significant, respectively. CV: Coefficient of variation. Traits include chlorophyll index (CI), leaf water potential (LWP), relative growth rate (RGR), relative water content (RWC), chlorophylla (COLa), chlorophyllb (COLb), total chlorophyll (COLto), carotenoid (CAR), grain yield (GYD), net photosynthesis rate (NAR) and leaf surface index (LAI).

Table 4 Goodness-of-fit criteria in algorithms for predicting grain yield.

Area	ML criterion	RF		SVM		XGBoost	
		Train	Test	Train	Test	Train	Test
Karaj	R2	0.526	0.518	0.591	0.590	0.648	0.652
	RMSE	0.282	0.176	0.178	0.098	0.087	0.084
	MAPE	5.946	5.394	4.864	3.845	3.127	3.014
	MAD	0.136	0.133	0.130	0.162	0.124	0.109
Qazvin	R2	0.625	0.436	0.534	0.496	0.680	0.628
	RMSE	0.201	0.167	0.246	0.186	0.053	0.067
	MAPE	7.595	6.739	3.742	3.581	2.650	2.503
	MAD	0.162	0.142	0.153	0.158	0.789	0.740
Damavand	R2	0.562	0.386	0.680	0.639	0.745	0.698
	RMSE	0.271	0.213	0.190	0.152	0.094	0.114
	MAPE	4.598	4.201	3.422	3.941	1.782	2.046
	MAD	0.126	0.111	0.056	0.059	0.024	0.028

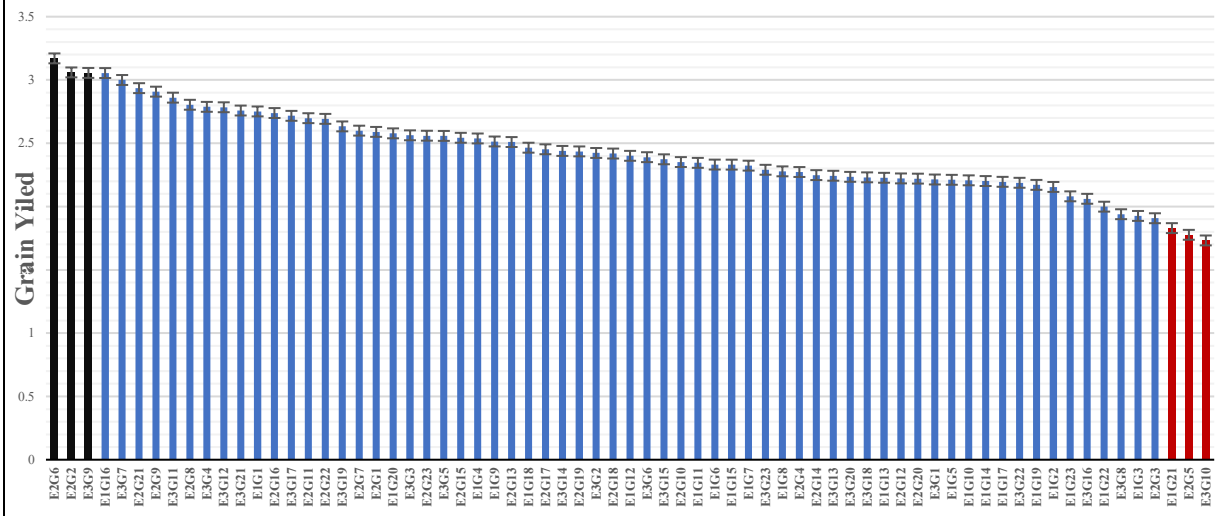

Fig. 1 Comparison of the average effect of genotype × environment in terms of grain yield traits on the genotypes investigated in the experiment E1: Karaj, E2: Qazvin, E3: Damavand.

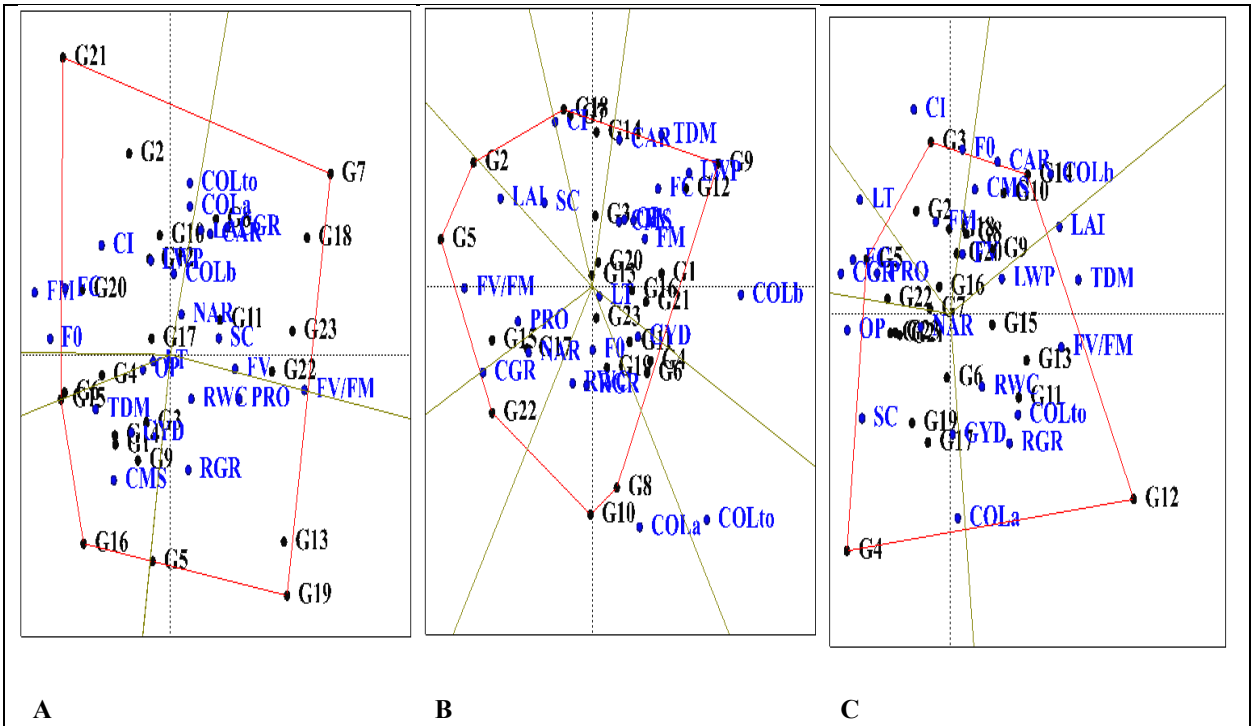

Fig. 2 Multivariate biplot of genotypes in terms of traits evaluated in the experiment, A: Karaj region, B: Damavand region, C: Qazvin region. Traits include product proline (Pro), chlorophylla (COLa), chlorophyllb (COLb), total chlorophyll (COLto), carotenoid (CAR), growth rate (CGR), dry matter production (TDM), osmotic potential (OP), cell membrane stability (CMS), leaf temperature (LT), leaf stomatal conductance (SC), florescence of chlorophyll (FC), fluorescence min (F0), fluorescence max (FM), variable fluorescence (FV), the ratio of variable fluorescence to maximum fluorescence (maximum quantum efficiency of photosystem II) (FV/FM), chlorophyll index (CI), leaf water potential (LWP), relative growth rate (RGR), relative water content (RWC), grain yield (GYD), net photosynthesis rate (NAR) and leaf surface index (LAI).

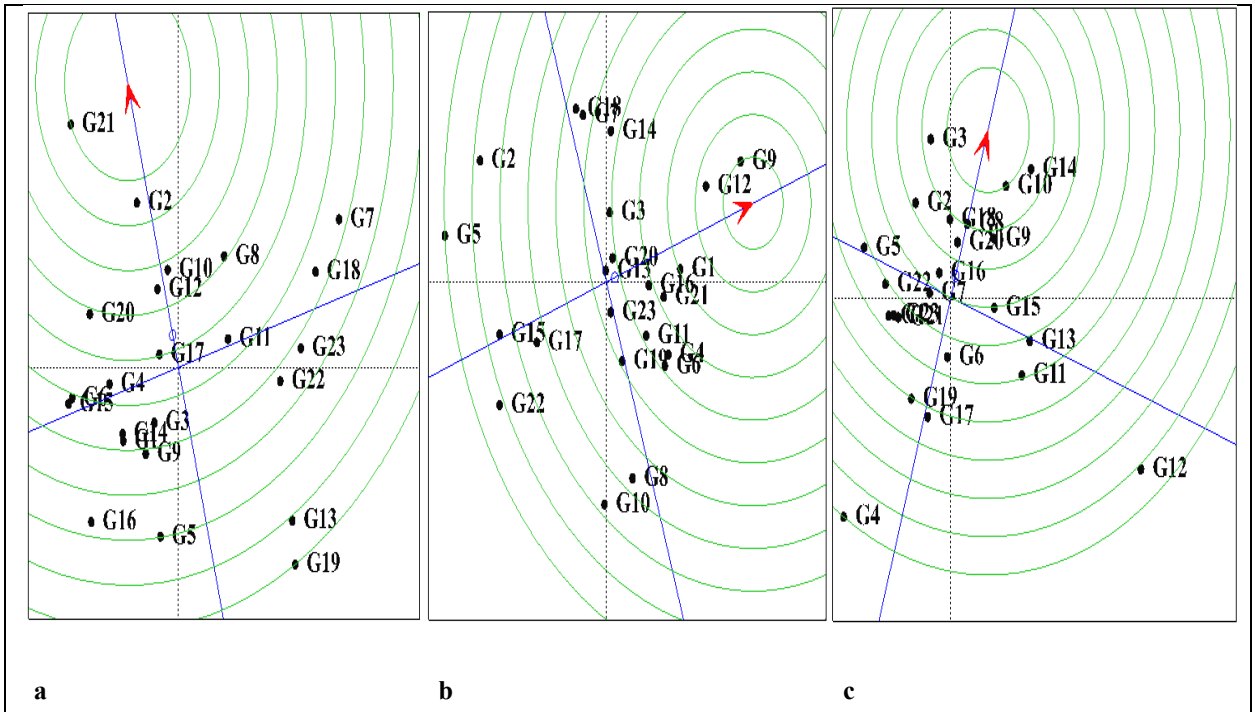

Fig. 3 Biplot ranking of genotypes based on the ideal genotype, a: Karaj region, b: Damavand region, c: Qazvin region

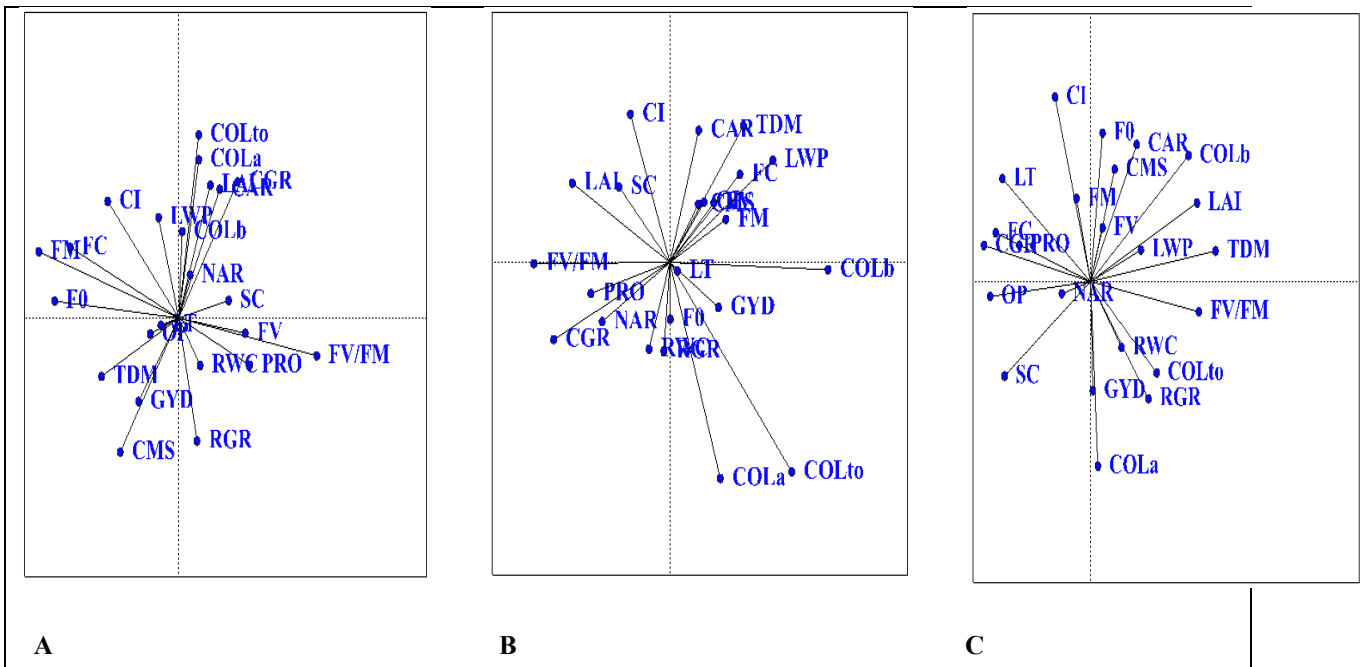

Fig. 4 Correlation diagram illustrating the relationships between the traits evaluated in the experiment. (A) Karaj region, (B) Damavand region, and (C) Qazvin region. The analyzed traits include proline content (Pro), chlorophyll a (COLa), chlorophyll b (COLb), total chlorophyll (COLto), carotenoids (CAR), crop growth rate (CGR), total dry matter production (TDM), osmotic potential (OP), cell membrane stability (CMS), leaf temperature (LT), stomatal conductance (SC), chlorophyll fluorescence (FC), minimum fluorescence (F_0), maximum fluorescence (Fm), variable fluorescence (F_v), the ratio of variable to maximum fluorescence (maximum quantum efficiency of photosystem II, F_v/Fm), chlorophyll index (CI), leaf water potential (LWP), relative growth rate (RGR), relative water content (RWC), grain yield (GYD), net assimilation rate (NAR), and leaf area index (LAI).

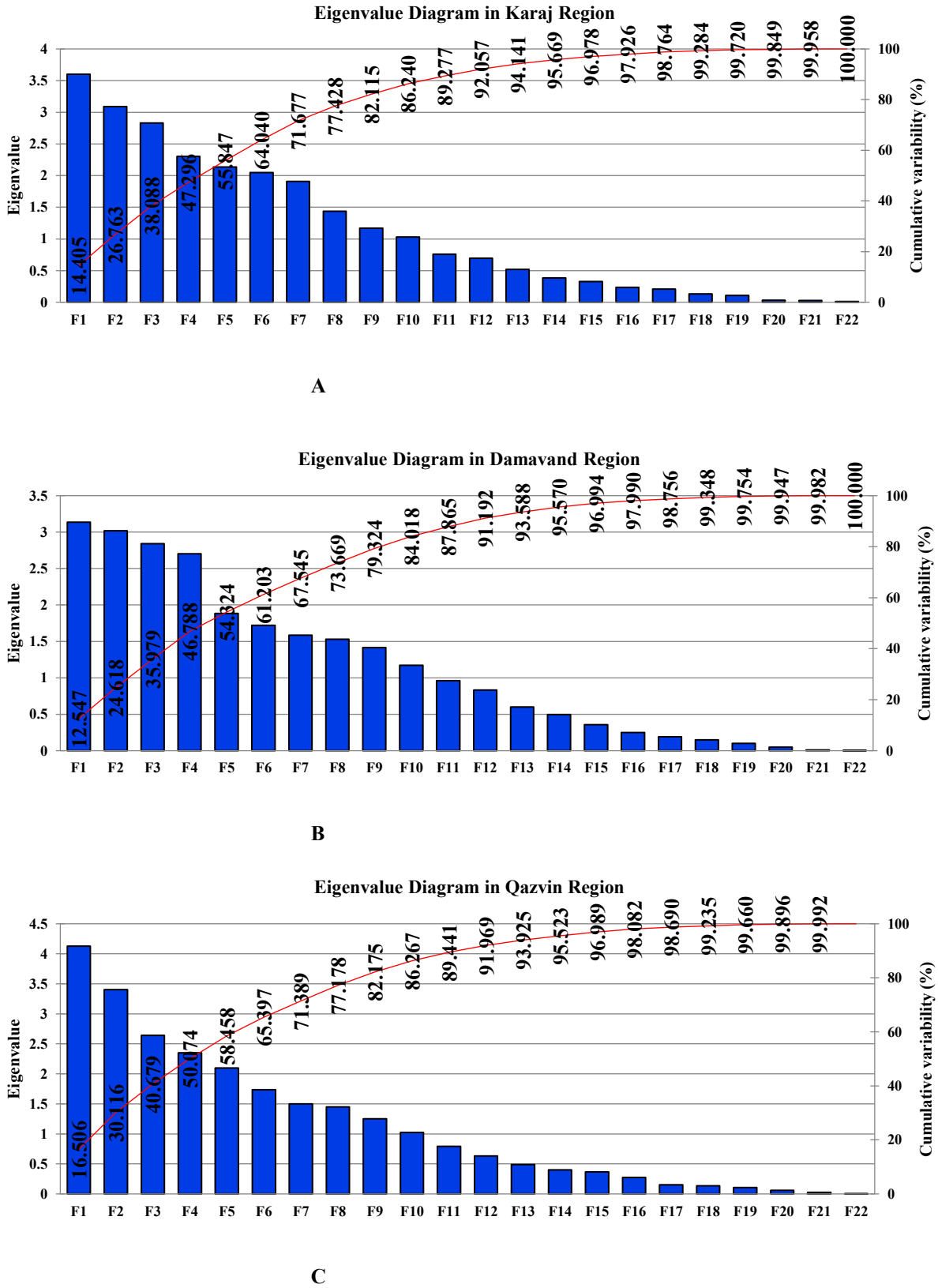

Fig. 5 Eigenvalue diagram of principal components decomposition in the logic examined in the experiment. A: Karaj region, B: Damavand region, C: Qazvin region

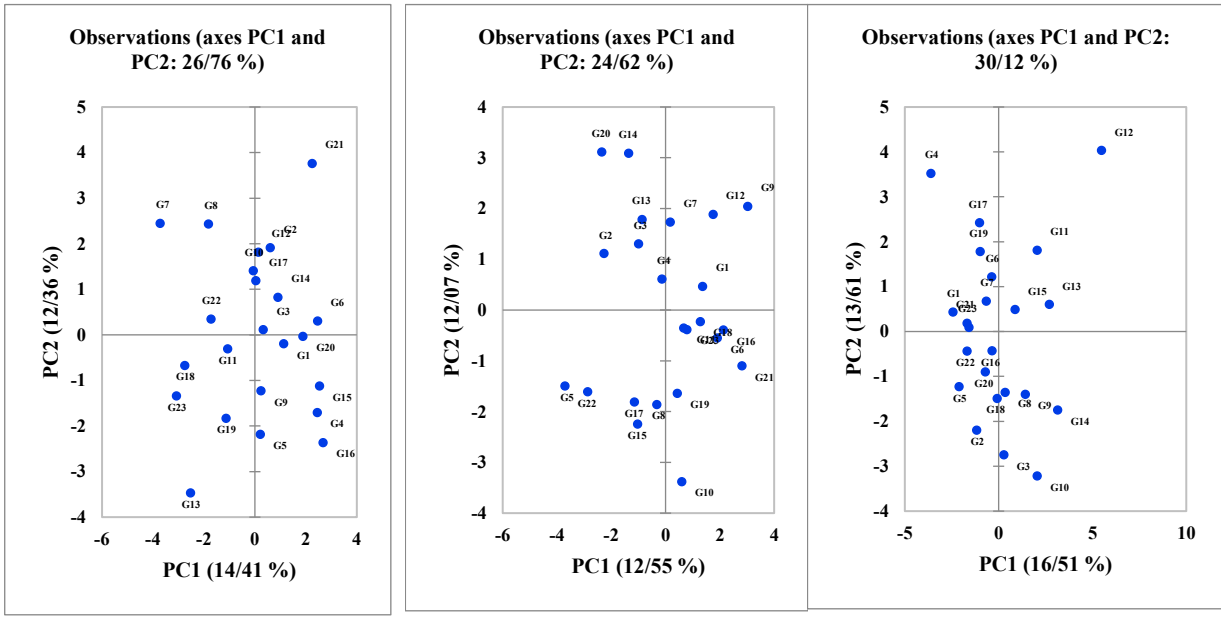

Fig. 6 The distribution diagram of the genotypes evaluated in the experiment in terms of the first principal component (PC1) and the second principal component (PC2), A: Karaj region, B: Damavand region, C: Qazvin region

PC1 and PC2 together explain approximately 26–30% of the total variation. Based on trait correlations and biplot patterns, PC1 appears to be primarily associated with biomass accumulation and photosynthetic efficiency, while PC2 is more closely related to plant water status and stress response. The exact biological interpretation of each component may vary slightly across environments due to genotype × environment interactions.