Prediction of Quinoa Grain Yield Using Machine Learning Models Based on Growth Traits and Yield Components under Deficit Irrigation and Organic Fertilization

Entessar Al-Jbawi*(1), Rasha Dannoura(2), and Abdullah Yacoub(3)

(1). Sugar Beet Research Department, Crops Research Administration, General Commission for Scientific Agricultural Research (GCSAR), Damascus, Syria.
(2). (GCSAR), Damascus, Syria.
(3). Faculty of Agriculture, Department of Rural Engineering, Damascus University, Damascus, Syria.
(*Corresponding author: dr.entessara@gmail.com or dr.entessara@gcsar.gov.sy).

Received: 01/05/2026 Accepted: 15/06/2026

Abstract

Machine learning (ML) techniques have become promising tools for predicting crop productivity due to their ability to analyze complex relationships between agronomic traits and yield performance and to improve prediction accuracy. However, studies comparing different machine learning models for predicting quinoa (*Chenopodium quinoa* Willd.) grain yield under deficit irrigation and organic fertilization conditions remain limited, particularly when dealing with small experimental datasets. This study aimed to compare the performance of four machine learning models, namely Linear Regression (LR), Random Forest (RF), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost), for predicting quinoa grain yield based on agronomic traits and yield components. The study was based on data obtained from a factorial field experiment consisting of 18 observations. Prior to model development, exploratory data analysis, correlation analysis, and feature selection using the Random Forest model were performed to identify the most influential variables affecting grain yield prediction. Model performance was evaluated using Leave-One-Out Cross-Validation (LOOCV), based on the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). In addition, the Wilcoxon Signed-Rank Test was applied to determine the statistical significance of differences among prediction errors of the evaluated models. Feature importance analysis revealed that grain weight per plant was the most influential factor for predicting grain yield, followed by organic fertilization, dry matter production, harvest index, and plant density. The results showed that the Linear Regression model achieved the best predictive performance among all evaluated models, with a coefficient of determination of: R^2=0.9968, while the values of root mean square error and mean absolute error were: RMSE=0.0176, MAE=0.0142. The Linear Regression model clearly outperformed the XGBoost, Random Forest, and

Support Vector Regression models. Furthermore, the Wilcoxon Signed-Rank Test confirmed that the differences in prediction errors between the Linear Regression model and the other models were statistically significant ($P < 0.05$). These findings indicate that simple statistical models may outperform more complex machine learning approaches when analyzing small experimental datasets characterized by strong linear relationships among variables. The study also provides an integrated analytical framework combining feature selection, cross-validation, and statistical comparison of machine learning models, contributing to improved crop yield prediction accuracy and offering potential applications in plant breeding programs and similar agricultural experiments.

Key words: Quinoa; Machine learning; Grain yield prediction; Linear Regression; Random Forest; Support Vector Regression; XGBoost; Feature selection; Leave-One-Out Cross-Validation; Deficit irrigation; Organic fertilization.

1-Introduction:

Quinoa (*Chenopodium quinoa* Willd.) is considered one of the most promising crops that has received increasing global attention during the last two decades due to its high nutritional value and exceptional ability to adapt to marginal environments. Its grains contain high-quality proteins rich in essential amino acids, in addition to considerable amounts of dietary fiber, minerals, vitamins, and antioxidants. These characteristics have positioned quinoa as a strategic crop that can contribute to enhancing global food security. Besides its nutritional importance, quinoa is characterized by its ability to grow under drought, salinity, and variable temperature conditions, making it one of the most suitable crops for cultivation in arid and semi-arid regions suffering from limited water resources and declining soil fertility. With the increasing impacts of climate change and growing pressure on natural resources, expanding quinoa cultivation has become a promising option for achieving more sustainable agricultural production. Consequently, many countries have incorporated quinoa into their research and development programs and have focused on improving its agronomic practices to enhance productivity and resource-use efficiency.

Although quinoa is considered one of the most tolerant crops to environmental stresses compared with many cereal crops, its productivity is still considerably affected by agronomic management practices, particularly irrigation and fertilization (Jacobsen et al., 2003; Bazile et al., 2016). In arid and semi-arid regions, water scarcity represents one of the major limiting factors for agricultural production. Therefore, increasing attention has been directed toward deficit irrigation (DI) strategies as an effective approach to improving water-use efficiency while maintaining acceptable crop productivity (Geerts and Raes, 2009; Hirich et al., 2014b). However, the success of this strategy depends on determining the appropriate level of irrigation reduction without causing significant reductions in plant growth and grain yield (Hirich et al., 2014b).

Several studies have demonstrated that quinoa responds differently to irrigation regimes depending on genotype and environmental conditions. Reducing irrigation water generally results in decreased values of some growth and yield-related traits; however, it may improve water-use efficiency under certain conditions (Ruiz et al., 2016; Hirich et al., 2014c).

On the other hand, organic fertilization is considered one of the most important agricultural practices contributing to the improvement of soil physical, chemical, and biological properties, increasing soil capacity for water and nutrient retention, and consequently reducing the adverse effects of water stress (Hirich et al., 2014a; Bazile et al., 2016). Previous studies have shown that the application of organic fertilizers or compost under deficit irrigation conditions improves quinoa growth and productivity and enhances water-use efficiency compared with deficit irrigation alone (Hirich et al., 2014a).

Despite considerable progress in understanding quinoa responses to irrigation and organic fertilization treatments, most previous studies have evaluated these effects using conventional statistical approaches, such as analysis of variance (ANOVA) and mean comparison tests, mainly to interpret differences among experimental treatments. However, these datasets have not been sufficiently exploited for developing accurate predictive models based on growth traits and yield components. This highlights the need for applying advanced analytical approaches, such as machine learning (ML) techniques, to develop predictive models capable of accurately estimating grain yield and supporting decision-making in quinoa crop management (Liakos et al., 2018; van Klompenburg et al., 2020; Shahhosseini et al., 2021).

In recent years, artificial intelligence applications, particularly machine learning techniques, have rapidly expanded in various fields of agricultural sciences due to their ability to analyze complex relationships among variables and identify hidden patterns that may be difficult to represent using traditional statistical methods (Khatibi and Ali, 2024; Jabed and Murad, 2024). Crop yield prediction has become one of the most important applications of machine learning, as it provides valuable information to researchers, decision-makers, and farmers for improving agricultural resource management, optimizing production practices, and strengthening food security under changing climatic conditions and increasing pressure on natural resources (Jabed and Murad, 2024).

A wide range of machine learning algorithms have been applied for crop yield prediction, including Linear Regression (LR), Random Forest (RF), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost). These models differ in their ability to represent linear and nonlinear relationships among variables (Khatibi and Ali, 2024; Govaichelvan et al., 2024). Linear Regression is commonly used as a baseline model due to its simplicity and interpretability, whereas Random Forest and XGBoost models are characterized by their ability to capture complex relationships and interactions among variables. Meanwhile, SVR has demonstrated good performance when dealing with small-sized or high-dimensional datasets (Khatibi and Ali, 2024).

Despite the extensive expansion of machine learning applications in crop yield prediction, most studies have relied mainly on climatic data, satellite imagery, or remote sensing information. In contrast, studies based on direct field measurements of growth traits and yield components, particularly in quinoa, remain relatively limited (Jabed and Murad, 2024). Furthermore, direct comparisons among Linear Regression, Random Forest, SVR, and XGBoost models using experimental field datasets under deficit irrigation and organic fertilization conditions are still scarce, emphasizing the need for further research in this area.

Although machine learning applications in crop yield prediction have advanced considerably, most published studies have primarily depended on meteorological data, soil characteristics, remote sensing indices, satellite images, or unmanned aerial vehicle (UAV) imagery. Meanwhile, direct field-based measurements, particularly growth traits and yield components, have received comparatively less attention in developing yield prediction models. Moreover, most studies have focused on major crops such as wheat, maize, rice, and soybean, whereas machine learning applications in quinoa remain limited, especially those integrating the effects of deficit irrigation and organic fertilization within a unified predictive framework.

Furthermore, many studies have focused on evaluating the performance of a single machine learning model or comparing a limited number of models, without conducting comprehensive comparisons among Linear Regression, Random Forest, Support Vector Regression, and XGBoost using a single experimental dataset. In addition, few studies have applied Leave-One-Out Cross-Validation (LOOCV), which is particularly suitable for small datasets, combined with statistical tests to verify the significance of differences among prediction errors.

Based on this research gap, the present study aimed to develop and compare four machine learning models, namely Linear Regression, Random Forest, Support Vector Regression (SVR), and XGBoost, for predicting quinoa grain yield based on growth traits and yield components under deficit irrigation and organic fertilization conditions. The study also aimed to evaluate model performance using LOOCV and statistical performance indicators, including R^2, RMSE, and MAE, in addition to applying the Wilcoxon Signed-Rank Test to determine the statistical significance of differences among prediction errors. This approach enables the identification of the most accurate and reliable model for quinoa grain yield prediction.

2. Materials and Methods

2.1. Study Site and Experimental Design

This study was based on data obtained from a field experiment conducted during the 2017 growing season at Qarhata Agricultural Research Station, affiliated with the Agricultural Scientific Research Center of Damascus, General Commission for Scientific Agricultural Research (GCSAR), Syria. The experimental site is located at 33.39° N latitude and 36.45° E longitude.

The experiment aimed to investigate the effects of different levels of deficit irrigation and organic fertilization on the growth and productivity of quinoa (Chenopodium quinoa Willd.). The data generated from this experiment were used as the basis for developing and evaluating machine learning models for grain yield prediction using measured agronomic and yield-related traits.

The soil of the experimental site was classified as sandy loam, with an alkaline reaction (pH = 8.0) and low electrical conductivity (ECe = 0.9 dS m^{-1}). The available contents of nitrogen, phosphorus, and potassium were 38.7, 7.6, and 62 mg kg^{-1}, respectively.

The experiment was conducted using a Randomized Complete Block Design (RCBD) arranged in a split-plot design with three replications. The main plots were assigned to three irrigation levels

representing 100%, 80%, and 60% of the crop water requirement, whereas the sub-plots were assigned to two organic fertilization treatments: control (without organic fertilization) and seaweed extract application.

2.2. Dataset Preparation

The study dataset was derived from a field experiment consisting of 18 experimental units resulting from the combination of three irrigation levels (100, 80, and 60% of full crop water requirement) and two organic fertilization treatments (without organic fertilization and seaweed extract) with three replications. Each experimental unit was considered an independent observation in the dataset used for developing machine learning models.

After completing the field experiment, all agronomic and yield-related measurements were collected for each experimental unit and organized into a digital database. Each row represented one experimental unit, whereas columns represented the investigated variables.

The study included a set of independent variables (predictor variables) comprising:

- Irrigation level (%)
- Organic fertilization treatment
- Plant density (plants ha^{-1})
- Dry matter yield (DMY)
- Plant weight (g)
- Grain weight per plant (g)
- Harvest index (%)
- Plant height (cm)
- Stem diameter (cm)
- Number of panicles per plant
- Panicle length (cm)
- Panicle diameter (cm)
- Panicle weight (g)
- 1000-grain weight (g)

Grain yield was considered the target variable, which the machine learning models aimed to predict.

Before model development, the dataset underwent an initial quality assessment, including checking for missing values, detecting outliers, verifying the absence of duplicated records, and assessing logical consistency among different variables. Due to the small dataset size and the absence of missing values, no records were removed or imputed.

2.3. Data Preprocessing

Data preprocessing represents a fundamental step in machine learning model development, as it aims to improve data quality and reduce potential sources of error that may affect predictive model performance (Géron, 2022).

In this study, the organic fertilization treatment was converted into a binary numerical variable (binary encoding), where the control treatment was assigned the value 0, while the seaweed extract treatment was assigned the value 1. Irrigation levels (60, 80, and 100%) were retained as their original numerical values because they represent quantitative levels corresponding to applied irrigation water percentages.

Descriptive statistical analysis was performed for all variables. In addition, a Pearson correlation matrix was generated to investigate preliminary relationships between agronomic traits and grain yield and to identify highly correlated independent variables.

Outliers were examined using boxplot analysis, and the dataset was rechecked to confirm the absence of missing values or duplicated observations before model training.

Since the dataset was complete, neither data imputation methods nor deletion of observations were required.

Feature scaling was not applied because all models were trained using the same dataset to ensure a fair comparison among algorithms. Furthermore, Linear Regression, Random Forest, and XGBoost models are not directly dependent on variable scaling, whereas scaling requirements were considered during the preparation of the Support Vector Regression (SVR) model.

2.4. Exploratory Data Analysis (EDA)

Before developing machine learning models, an Exploratory Data Analysis (EDA) was performed to understand the characteristics of the dataset and identify the nature of relationships among the studied variables, which represents an essential step prior to predictive modeling (James et al., 2021).

The exploratory analysis included calculating descriptive statistics for all variables, including the mean, standard deviation, minimum, and maximum values, as well as visualizing data distributions using appropriate graphical methods.

A Pearson correlation matrix was also constructed to evaluate the strength and direction of relationships between predictor variables and grain yield. Pearson’s correlation coefficient was calculated according to the following equation:

The correlation coefficient ranges between −1 and +1, where positive values indicate a direct relationship, negative values indicate an inverse relationship, and values close to zero indicate a weak linear relationship between the two variables (Montgomery et al., 2021).

This analysis contributed to identifying the variables most strongly associated with grain yield and provided an initial understanding of the internal relationships among agronomic traits before proceeding to the development of predictive models.

2.5. Feature Selection

Feature selection is an essential step in machine learning model development, as it helps identify the variables that contribute most to the prediction of the target variable, improves model interpretability, reduces computational complexity, and minimizes the risk of overfitting (Kuhn & Johnson, 2019).

In this study, all measured agronomic traits were considered as independent variables (features), whereas grain yield was defined as the target variable. No variables were excluded prior to model training because of the relatively small size of the dataset. Instead, the relative importance of each predictor was assessed after training the Random Forest (RF) model.

Feature importance was estimated using the Random Forest Feature Importance method, which quantifies the contribution of each predictor based on the total reduction in node impurity achieved when that feature is used for splitting during the construction of decision trees. The contribution of each feature is then averaged across all trees in the forest (Breiman, 2001). The overall importance of the j-th feature was calculated as:

The calculated feature importance scores were subsequently used to rank the predictor variables in descending order according to their contribution to explaining the variation in grain yield. The results were visualized using a Feature Importance Plot, providing a clear representation of the relative importance of each agronomic trait.

2.6. Machine Learning Models

The study aimed to compare the predictive performance of four regression models representing different levels of mathematical complexity for quinoa grain yield prediction, ranging from a conventional statistical model to advanced nonlinear machine learning algorithms.

2.6.1. Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) was employed as the baseline model to describe the linear relationship between the predictor variables and grain yield. The model is expressed as follows:

The regression coefficients were estimated using the Ordinary Least Squares (OLS) method, which minimizes the Residual Sum of Squares (RSS) between the observed and predicted values (James et al., 2021).

2.6.2. Random Forest Regression (RFR)

Random Forest Regression (RFR) is a supervised ensemble learning algorithm that combines the predictions of multiple decision trees to improve predictive accuracy and reduce model variance. Each decision tree is trained using a bootstrap sample of the training data, while a random subset of predictor variables is considered at each split, thereby increasing model diversity and reducing the risk of overfitting (Breiman, 2001).

The final prediction is obtained by averaging the predictions of all individual decision trees, as expressed by:

The Random Forest model implemented in this study consisted of 200 decision trees (n_estimators = 200), a value selected to provide a suitable balance between prediction accuracy, model stability, and computational efficiency. A fixed random seed (random_state = 42) was used to ensure the reproducibility of the results (Breiman, 2001).

2.6.3. Support Vector Regression (SVR)

Support Vector Regression (SVR) was employed because of its ability to model both linear and nonlinear relationships, particularly when dealing with relatively small datasets. Unlike conventional regression methods, SVR seeks to identify an optimal regression function that minimizes prediction error while maximizing the margin around the fitted function, thereby improving the model's generalization capability (Smola & Schölkopf, 2004).

The SVR model is expressed as:

In this study, the Radial Basis Function (RBF) kernel was employed because of its strong capability to capture complex nonlinear relationships between predictor variables and grain yield (Smola & Schölkopf, 2004

2.6.4. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) was employed as one of the most advanced machine learning algorithms for regression tasks. The algorithm is based on a sequential ensemble learning approach in which decision trees are built iteratively, with each new tree trained to correct the prediction errors of the previously constructed trees (Chen & Guestrin, 2016).

The XGBoost model can be expressed mathematically as:

XGBoost is characterized by its ability to model complex nonlinear relationships while achieving high predictive accuracy. Furthermore, it incorporates regularization techniques that improve model generalization and reduce the risk of overfitting, making it particularly effective for regression and predictive modeling tasks (Chen & Guestrin, 2016).

2.7. Model Training and Validation

Evaluating the ability of a predictive model to generalize to unseen data is a fundamental step in machine learning model development. The primary objective is to assess the model's generalization performance, ensuring that its predictive capability extends beyond the training data.

Because the dataset used in this study was relatively small (18 observations), Leave-One-Out Cross-Validation (LOOCV) was adopted as the validation strategy. LOOCV is considered one of the most appropriate validation methods for small datasets because it maximizes the use of available observations for model training while providing an unbiased estimate of predictive performance (Stone, 1974).

In LOOCV, the model is trained using 17 observations, while the remaining observation is reserved for testing. This procedure is repeated until every observation has served exactly once as the test sample. The predictions obtained from all iterations are then combined to calculate the overall performance metrics.

The LOOCV prediction error can be expressed as:

LOOCV reduces the bias associated with conventional train-test data partitioning by allowing nearly all observations to be used for model training in each iteration. Consequently, it provides a more reliable estimate of predictive performance when working with small datasets.

The same LOOCV procedure was applied to all machine learning models evaluated in this study to ensure a fair comparison of their predictive performance. Since the dataset contained 18 observations, a total of 18 validation iterations were performed, with 17 observations used for training and one observation used for testing in each iteration (Stone, 1974).

2.8. Model Evaluation

After model training, predictive performance was evaluated using three statistical metrics that are widely employed in regression analysis and agricultural prediction studies: the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE). These metrics provide complementary information regarding the goodness-of-fit and prediction accuracy of the evaluated models.

2.8.1. Coefficient of Determination (R^2)

The coefficient of determination (R^2) measures the proportion of the total variation in the target variable that is explained by the prediction model. It is calculated as follows:

The value of R^2 ranges from 0 to 1, with values closer to 1 indicating that the model explains a greater proportion of the variability in the observed data and therefore exhibits better predictive performance (James et al., 2021).

2.8.2. Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is used to quantify the average magnitude of prediction errors between the observed and predicted values. It is calculated according to the following equation (Willmott & Matsuura, 2005):

This metric gives greater weight to larger errors because the residuals are squared before averaging. Therefore, RMSE is considered a sensitive indicator for evaluating the accuracy and reliability of predictive models, particularly when large deviations between observed and predicted values occur.

2.8.3. Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) measures the average magnitude of the absolute differences between the observed and predicted values. It is calculated according to the following equation:

MAE is characterized by its straightforward interpretation and is less influenced by extreme values compared with RMSE, as it does not involve squaring the prediction errors (Willmott & Matsuura, 2005).

The three evaluation metrics (R^2, RMSE, and MAE) were used collectively to provide a comprehensive assessment of model performance. While R^2 reflects the explanatory ability of the model, RMSE and MAE evaluate prediction accuracy from two complementary perspectives, considering the magnitude of prediction errors in different ways.

2.9. Statistical Comparison of Models

The comparison among machine learning models was not limited to predictive performance metrics. In addition, the Wilcoxon Signed-Rank Test was applied to compare the absolute prediction errors of the best-performing model with those of the remaining models.

This test was selected because it is a non-parametric statistical test that does not require the data to follow a normal distribution. It is also appropriate for comparing two related samples with small sample sizes, which is consistent with the characteristics of the dataset used in this study (Wilcoxon, 1945).

The test was based on the differences between the absolute errors of the compared models. For each observation, the error difference was calculated as:

The absolute values of these differences were then calculated:

The absolute differences were ranked in ascending order, and each value was assigned a rank while retaining the original sign of the difference (positive or negative).

The Wilcoxon signed-rank statistic was then calculated as:

A significance level of 0.05 was adopted to determine whether statistically significant differences existed among model performances. A P-value < 0.05 was considered evidence of a significant difference between the prediction errors of the compared models.

2.10. Software Environment

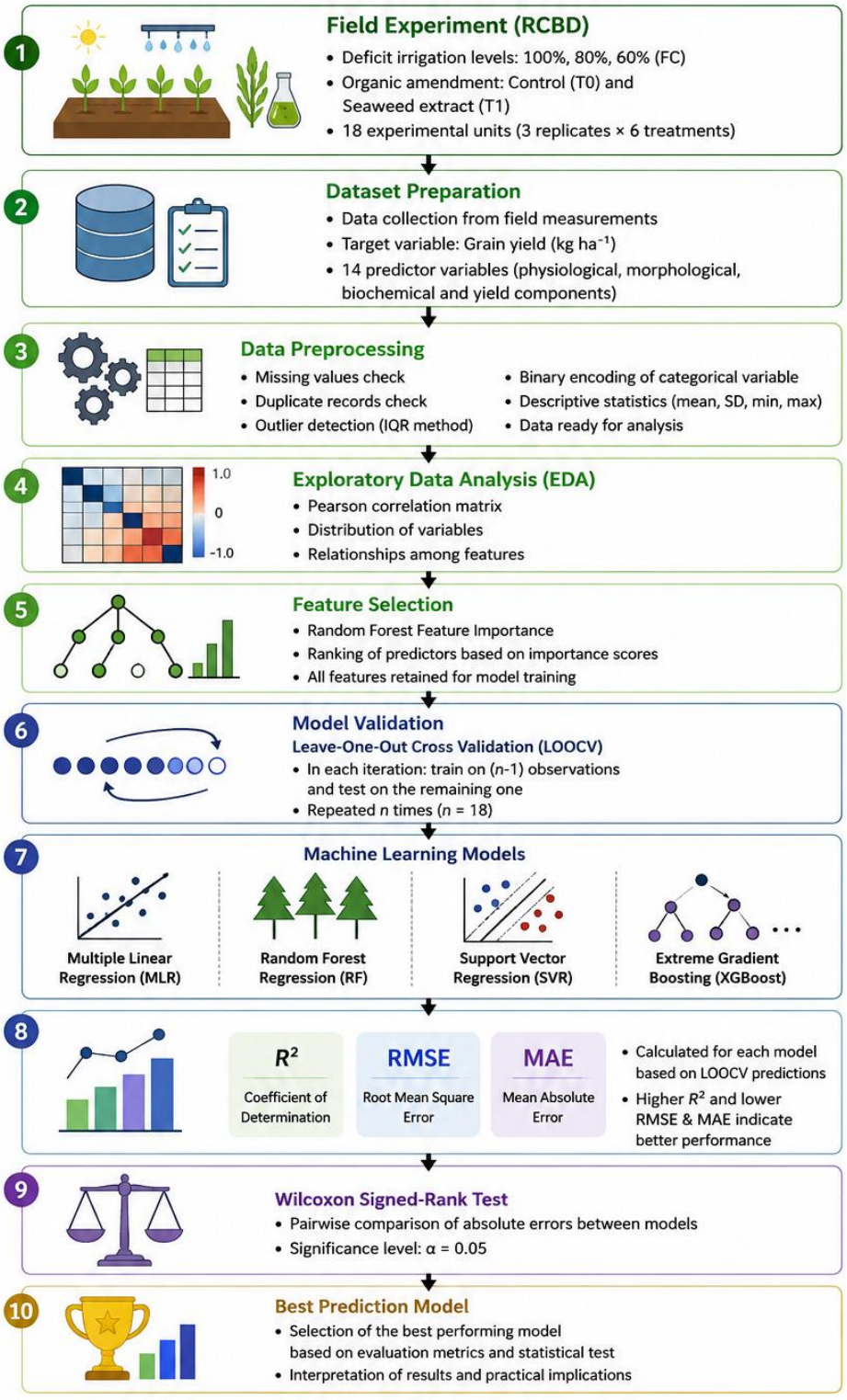
All data preprocessing, model development, cross-validation procedures, and calculation of statistical performance metrics were performed using the Python programming language (version 3.x).

A set of specialized scientific libraries was employed, including:

- Pandas for data management, manipulation, and preprocessing.
- NumPy for numerical and array-based computations.
- Scikit-learn for developing the Linear Regression, Random Forest Regression, and Support Vector Regression models, as well as implementing the Leave-One-Out Cross-Validation (LOOCV) procedure.
- XGBoost for developing the Extreme Gradient Boosting model.
- SciPy for performing the Wilcoxon Signed-Rank Test.
- Matplotlib and Seaborn for generating graphical visualizations, including the correlation matrix, feature importance plots, and observed versus predicted value relationships.

To ensure reproducibility of the results, a fixed random seed (random_state = 42) was applied to all stochastic models, and identical computational settings were maintained throughout all model development and comparison procedures.

Figure 1.
Workflow of the Machine Learning Framework for Predicting Quinoa Grain Yield



FC: Field Capacity; RCBD: Randomized Complete Block Design; LOOCV: Leave-One-Out Cross Validation; R^2: Coefficient of Determination; RMSE: Root Mean Square Error; MAE: Mean Absolute Error.

3. Results

3.1. Data Preparation and Exploratory Data Analysis

3.1.1. Dataset Organization and Descriptive Statistics

The initial dataset was prepared by entering all field measurements into **Microsoft Excel**, followed by variable encoding, data validation, and defining **grain yield** as the **target variable** for the machine learning models.

An **Exploratory Data Analysis (EDA)** was performed to characterize the statistical properties of the input variables and to identify the variability and preliminary relationships among them before developing the predictive models.

Table (1) presents the descriptive statistics of all variables used in this study, including the **mean, standard deviation (SD), minimum, and maximum values**.

Table (1). Descriptive statistics of the variables used for developing machine learning models.

Variable	Mean	Standard deviation	Minimum	Maximum
Irrigation (%)	80.00	16.80	60.00	100.00
Organic amendment	0.50	0.51	0	1
Plant number (1000 plants ha^{-1})	91.03	4.43	81.60	98.30
Dry matter yield (kg ha^{-1})	671.16	106.17	494.80	857.00
Plant weight (g)	67.01	9.89	50.00	85.00
Grain weight $plant^{-1}$ (g)	19.24	2.97	15.20	25.50
Grain yield (t ha^{-1})	1.75	0.32	1.30	2.30
Harvest index (%)	29.03	4.60	22.70	40.00
Plant length (cm)	101.78	13.90	80.00	130.00
Stem diameter (cm)	1.32	0.46	0.50	2.00
Panicle number $plant^{-1}$	19.72	4.62	12.00	28.00
Panicle length (cm)	39.89	6.08	30.00	50.00
Panicle diameter (cm)	9.06	2.16	5.00	12.00
Panicle weight (g)	36.67	7.90	21.50	47.30
Thousand grain weight (g)	1.61	0.19	1.31	2.00

The results showed considerable variation among observations for most measured traits, providing an appropriate range of variability required for training machine learning models. Grain yield ranged from **1.30 to 2.30 t ha^{-1}**, with an average value of **1.75 t ha^{-1}**, whereas dry matter yield ranged from **494.8 to 857.0 kg ha^{-1}**, with an average of **671.16 kg ha^{-1}**.

Morphological traits also exhibited clear differences among treatments. Plant length ranged from **80 to 130 cm**, while plant weight varied from **50 to 85 g**. The mean thousand grain weight was **1.61 g**. This variability indicates the presence of environmental and physiological differences among treatments, which can be effectively exploited for developing predictive models.

3.1.2. Analysis of Target Variable Distribution

The distribution of grain yield was analyzed using a **histogram** to evaluate the data distribution pattern and determine the spread of values within the dataset.

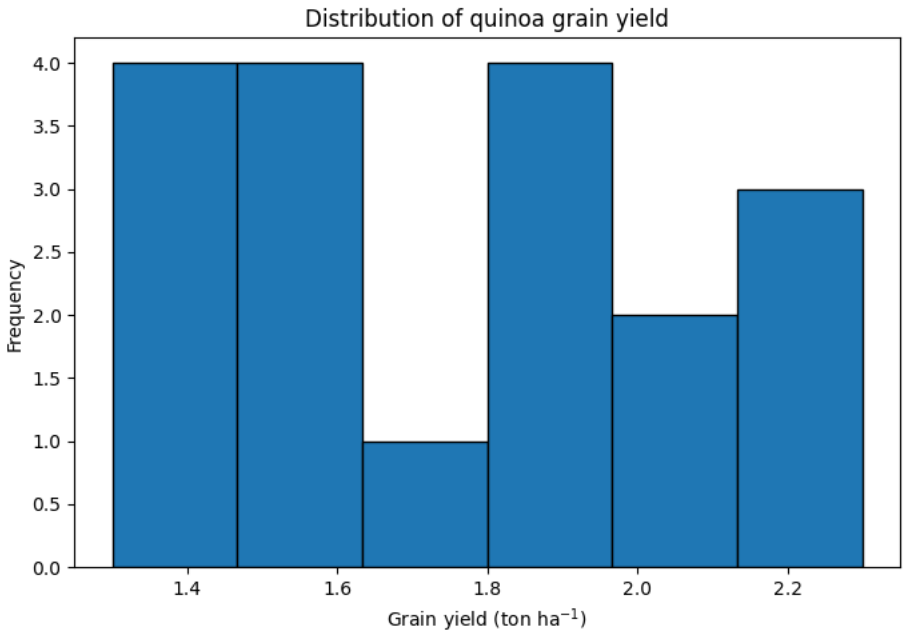


Figure (2). Distribution of quinoa grain yield values (t ha^{-1}) used for training and evaluating machine learning models.

The distribution of grain yield showed an appropriate spread of values within the experimental range, without abnormal clustering or severe deviations. This supports the suitability of the dataset for subsequent modeling procedures.

3.1.3. Correlation Analysis

The **correlation matrix** was used to determine the relationships between the input traits and grain yield, as well as to identify the interrelationships among predictor variables before developing the machine learning models.

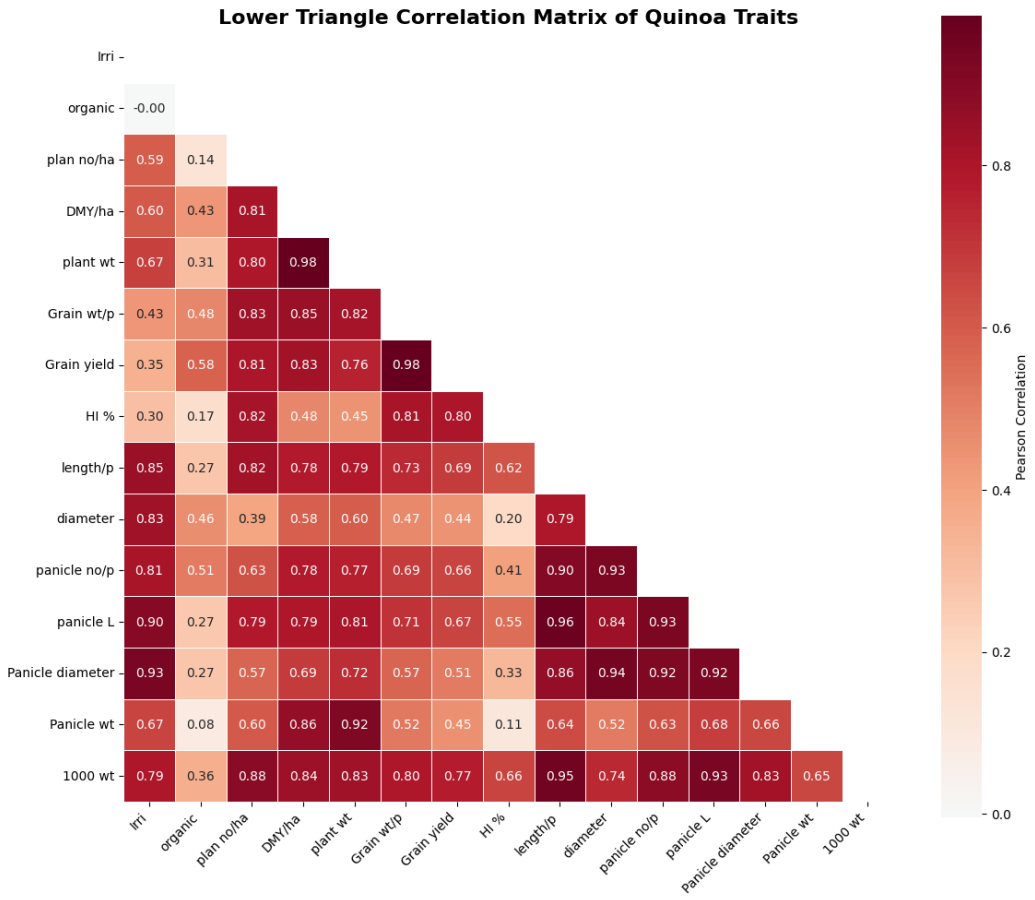


Figure (3). Correlation matrix among the variables used in the study.

The correlation matrix revealed different levels of association between growth traits, yield components, and grain yield of quinoa under different irrigation levels and organic amendment treatments.

Grain yield showed a **very strong positive correlation** with **grain weight per plant (Grain wt/p)**, with a correlation coefficient of:

$$r=0.974$$

indicating that increased productivity at the individual plant level was one of the most important determinants of final grain yield per unit area.

Grain yield also exhibited a strong positive correlation with **organic amendment**, with:

$$r=0.854$$

reflecting the positive role of organic inputs in improving plant growth and enhancing yield potential.

Furthermore, grain yield was positively correlated with **plant density (Plant number ha^{-1})**:

$$r=0.659$$

and with **dry matter yield (DMY)**:

$$r=0.599$$

indicating that increased vegetative growth and biomass accumulation contributed significantly to improved grain production.

The **harvest index (HI%)** showed a moderate positive correlation with grain yield:

$$r=0.542$$

suggesting the importance of efficient allocation of accumulated dry matter toward grain formation.

In contrast, irrigation level showed a weak negative correlation with grain yield:

$$r=-0.264$$

This does not indicate that reduced irrigation increased grain production; rather, it reflects the distribution pattern of irrigation treatments in the experiment and the interaction between irrigation level and organic amendment, as the relationship between applied water amount and grain yield was not a simple linear relationship.

The correlation matrix also revealed strong relationships among several morphological traits, particularly **panicle traits**, where correlation coefficients among panicle length, panicle diameter, and panicle number were relatively high:

$$r>0.80$$

These relationships indicate the presence of **multicollinearity** among predictor variables, which should be considered during feature selection and machine learning model development.

Table (2). Definitions of abbreviations and variables used for developing machine learning models for quinoa grain yield prediction.

Variable	Definition	Abbreviation
Irrigation	Irrigation level (60, 80, and 100%)	Irri
Organic amendment	Organic amendment (0 = without amendment, 1 = organic amendment)	Organic
Plant density	Number of plants per hectare	plan no/ha
Dry matter yield	Dry matter yield (kg ha^{-1})	DMY/ha
Plant weight	Plant weight (g)	plant wt
Grain weight $plant^{-1}$	Grain weight per plant (g)	Grain wt/p
Plant length	Plant length (cm)	length/p
Stem diameter	Stem diameter (cm)	diameter
Panicle number	Number of panicles per plant	panicle no/p
Panicle length	Panicle length (cm)	panicle L
Panicle diameter	Panicle diameter (cm)	Panicle diameter
Panicle weight	Panicle weight (g)	Panicle wt
Thousand grain weight	Thousand grain weight (g)	1000 wt
Harvest index	Harvest index (%)	HI %
Grain yield	Grain yield (t ha^{-1})	Grain yield

3.1.4. Outlier Detection

Boxplot analysis was performed for all variables to detect potential outliers. The results showed that none of the variables contained abnormal values according to the criterion:

$$1.5\times IQR$$

This indicates the homogeneity of the dataset and confirms that no observations required removal or adjustment before machine learning model development. Therefore, all available data were retained during the model training and evaluation procedures.

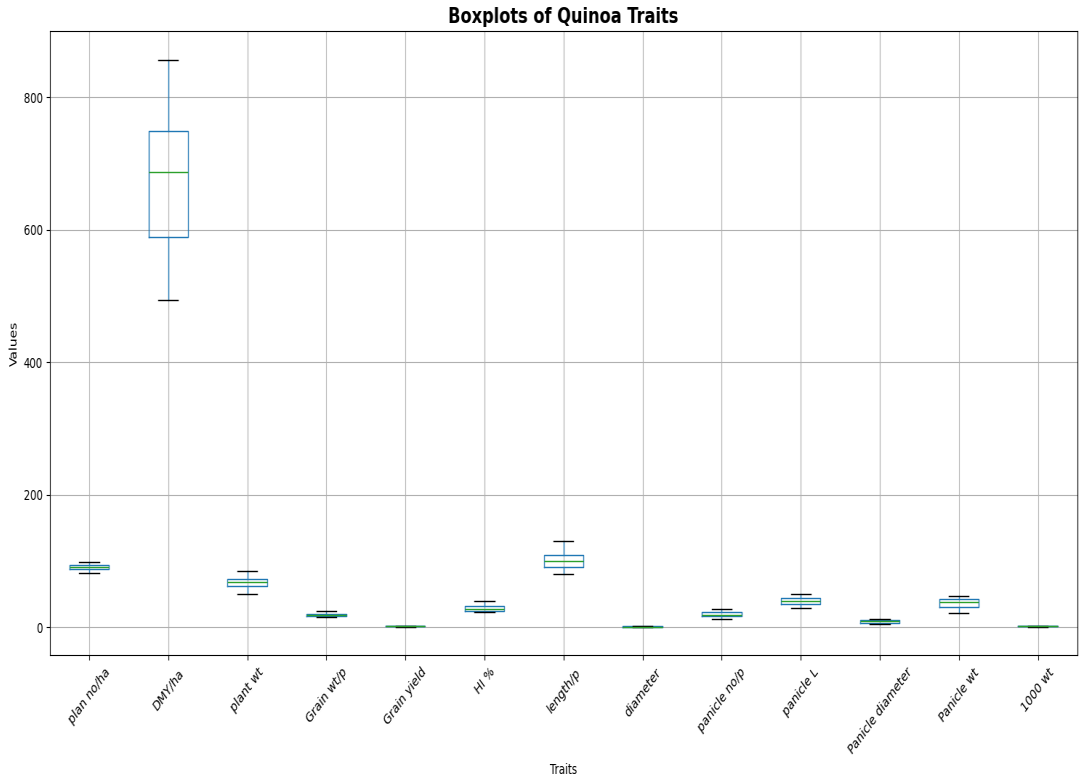


Figure (4). Boxplot analysis of all variables.

3.2. Feature Selection

The **Random Forest Regression (RFR)** model was used to determine the relative importance of independent variables in explaining the variation in quinoa grain yield.

Variables were ranked according to their **feature importance values**, and the most influential variables were identified to evaluate their contribution to grain yield prediction using machine learning models.

Table (3). Ranking of variables according to their importance using the Random Forest model.

Rank	Variable	Importance
1	Grain wt/p	0.723
2	Organic	0.058
3	DMY/ha	0.052
4	HI %	0.034
5	plan no/ha	0.031
6	panicle no/p	0.023
7	plant wt	0.019
8	Panicle wt	0.019
9	1000 wt	0.015
10	panicle L	0.008
11	diameter	0.007
12	length/p	0.005
13	Irri	0.003
14	Panicle diameter	0.002

The Random Forest model indicated that **grain weight per plant (Grain wt/p)** was the most important variable for explaining grain yield variation, with a feature importance value of **0.723**, followed by **organic amendment (0.058)** and **dry matter yield (DMY/ha) (0.052)**.

The remaining variables showed relatively lower importance values, indicating a smaller contribution to improving the model's predictive ability for grain yield.

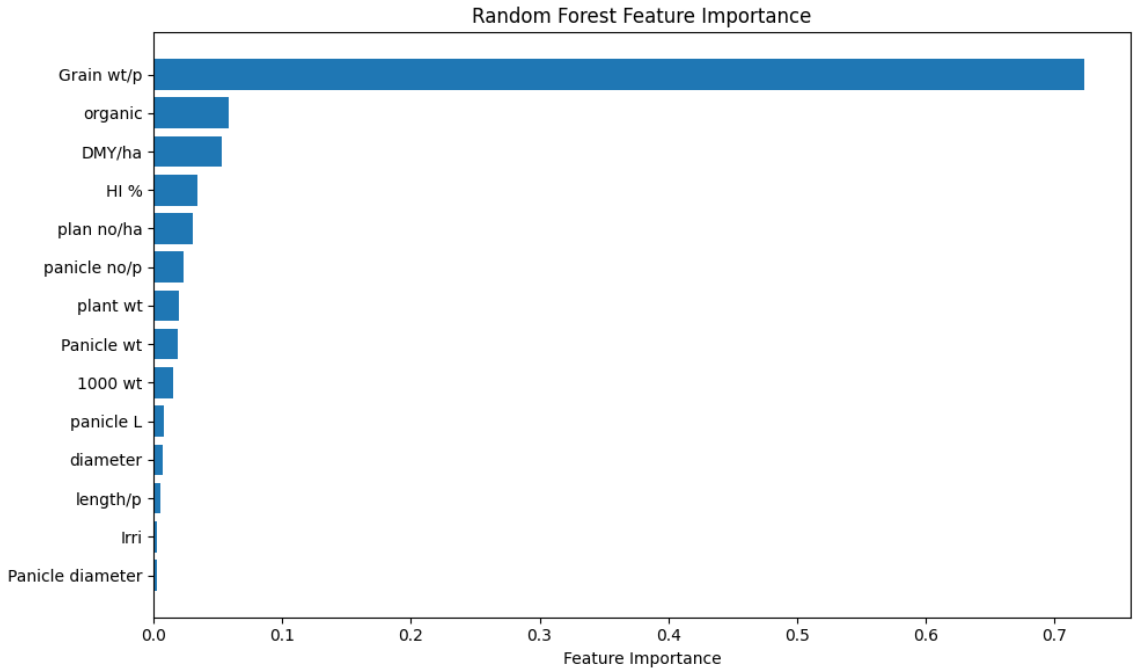


Figure (5). Feature importance based on the Random Forest model.

The feature importance ranking demonstrated that **grain weight per plant** was the dominant predictor of quinoa grain yield, with an importance value of **0.723**, which was considerably higher than all other variables. This result reflects the close relationship between individual plant grain productivity and final grain yield per hectare.

Organic amendment ranked second, with an importance value of **0.058**, followed by dry matter yield (**0.052**), indicating their contribution to improving the model's ability to explain yield variation.

Other variables, including harvest index (HI), plant density, panicle number, and plant weight, showed relatively lower importance values, whereas irrigation level and panicle diameter were the least influential variables.

The high importance of grain weight per plant was expected because it represents a direct yield component; therefore, its high contribution in the model reflects its strong relationship with the target variable.

3.3. Machine Learning Model Development and Evaluation

Due to the small size of the dataset (**18 observations**), the **Leave-One-Out Cross-Validation (LOOCV)** approach was adopted. In this procedure, each model was trained using **17 observations** and tested using the remaining observation, and this process was repeated until all observations had been used once as test samples.

This validation strategy is particularly suitable for small datasets because it maximizes the utilization of available data and reduces the bias associated with conventional data partitioning into training and testing subsets.

3.3.1. Performance of the Linear Regression Model

The **Linear Regression model** achieved the highest predictive performance among all evaluated models. The coefficient of determination reached:

R^2=0.9968

indicating that the model explained **99.68%** of the variation in quinoa grain yield.

Furthermore, the model recorded the lowest prediction errors:

RMSE=0.0176

and

MAE=0.0142

demonstrating its very high predictive accuracy.

3.3.2. Performance of the Random Forest Regression Model

The **Random Forest Regression model** showed good predictive capability, achieving:

R^2=0.8210

with:

RMSE=0.1316

and:

MAE=0.1049

These results indicate that the model was able to explain a substantial proportion of the variability in quinoa grain yield.

3.3.3. Performance of the Support Vector Regression (SVR) Model

The **Support Vector Regression (SVR)** model exhibited moderate performance compared with the other evaluated models, with:

R^2=0.6784

and prediction errors of:

RMSE=0.1765

and:

MAE=0.1399

Among the four evaluated models, SVR showed the lowest predictive accuracy.

3.3.4. Performance of the XGBoost Regression Model

The **XGBoost Regression model** achieved high predictive performance, with:

R^2=0.8403

and recorded:

RMSE=0.1244

and:

MAE=0.0848

These results demonstrate the ability of XGBoost to effectively capture nonlinear relationships between agronomic traits and quinoa grain yield.

3.4. Final Comparison among Machine Learning Models

Table (4) presents a comparison of the predictive performance of the four evaluated models using the performance metrics **R^2, RMSE, and MAE**.

Table (4). Performance comparison of machine learning models used for quinoa grain yield prediction using LOOCV.

Rank	Model	R^2	RMSE (t ha^{-1})	MAE (t ha^{-1})
1	Linear Regression	0.9968	0.0176	0.0142
2	XGBoost Regression	0.8403	0.1244	0.0848
3	Random Forest Regression	0.8210	0.1316	0.1049
4	Support Vector Regression (SVR)	0.6784	0.1765	0.1399

The **Linear Regression model** clearly outperformed all other evaluated models, achieving the highest **R^2 value** and the lowest **RMSE and MAE values**, followed by **XGBoost Regression** and **Random Forest Regression**, whereas **SVR** showed the lowest predictive performance.

These findings indicate that the relationship between agronomic variables and quinoa grain yield in this study was highly amenable to linear modeling, despite the evaluation of more complex machine learning algorithms.

3.5. Statistical Significance Test for Model Comparison

The **Wilcoxon Signed-Rank Test** was applied to compare the absolute prediction errors of the **Linear Regression model** with those of the other evaluated models.

Table (5). Wilcoxon Signed-Rank Test results comparing prediction errors between Linear Regression and the other models.

Comparison	Test statistic (W)	P-value	Result
Linear Regression vs. Random Forest	3.0	<0.001	Significant difference
Linear Regression vs. SVR	0.0	<0.001	Significant difference
Linear Regression vs. XGBoost	0.0	<0.001	Significant difference

Note: Statistical significance was determined at **$P < 0.05$**.

The results revealed highly significant differences:

P<0.001P<0.001P<0.001

between the prediction errors of the **Linear Regression model** and those of **Random Forest, SVR, and XGBoost**.

These findings indicate that the superior performance of the Linear Regression model was not due to random variation, but rather reflected a true improvement in predictive capability under the conditions of this study.

3.6. Visual Evaluation of Model Performance

Figure (6) illustrates the comparison among the evaluated models based on different performance metrics.

The graphical representation confirmed the superiority of the **Linear Regression model**, which achieved the highest coefficient of determination and the lowest prediction errors. Meanwhile, **XGBoost and Random Forest Regression** showed relatively similar performance levels, whereas **SVR exhibited the lowest predictive capability**.

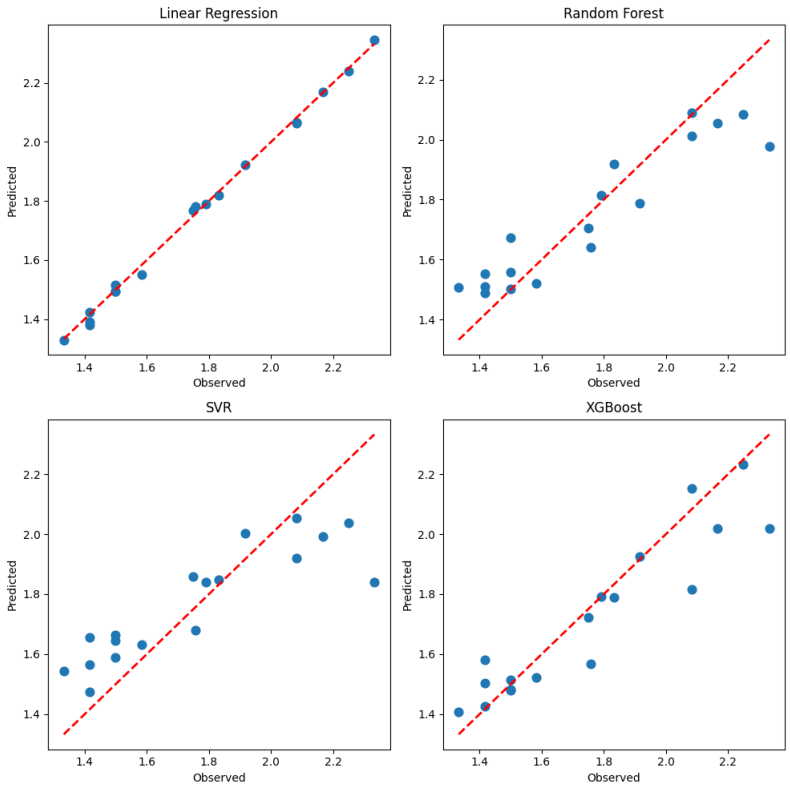


Figure (6). Comparison of machine learning models based on different performance metrics.

4. Discussion

4.1. Importance of Variables Affecting Quinoa Grain Yield

Feature importance analysis using the Random Forest model revealed that grain weight per plant was the most influential variable in predicting quinoa grain yield, with an importance value of 0.723, exceeding all other variables by a large margin. Organic amendment and dry matter yield ranked second and third, respectively.

This finding is physiologically reasonable because grain yield per hectare represents the direct outcome of grain production per individual plant combined with plant density per unit area. Therefore, any increase in grain weight per plant is directly reflected in the final grain yield. Bazile et al. (2016) indicated that grain weight per plant is among the most stable yield components in quinoa and is considered a key indicator in breeding programs aimed at improving productivity.

Organic amendment ranked second in importance, highlighting its role in improving soil physical, chemical, and biological properties, enhancing water retention capacity and nutrient availability, particularly under deficit irrigation conditions. This result agrees with the findings of Mirsafi et al. (2024a,b), who demonstrated that improved water management combined with appropriate agronomic practices enhances quinoa growth, water use efficiency, and final yield.

The third rank of dry matter yield (DMY) indicates that biomass production capacity remains strongly associated with grain yield, as dry matter provides the basis for carbohydrate accumulation and translocation toward seeds during grain filling. Several studies have confirmed that dry matter production is one of the most reliable indirect indicators for estimating grain yield in quinoa and other field crops (Geerts & Raes, 2009; Awa et al., 2024).

In contrast, some traits, such as panicle diameter and irrigation level, showed relatively low importance values in the Random Forest model. This does not indicate that these factors have no effect on grain yield; rather, it suggests that their effects may have been indirectly represented through other variables more closely associated with yield, such as grain weight per plant and dry matter yield. This behavior is common in machine learning models when correlations or interactions exist among predictor variables, as algorithms tend to assign greater importance to variables with higher explanatory power for the target variable (Breiman, 2001; Kuhn & Johnson, 2019).

These findings emphasize the importance of applying Feature Importance techniques in agricultural studies, as they not only improve the predictive performance of models but also help identify the most relevant traits that should be prioritized in crop breeding programs and crop management practices. Therefore, the results provide both predictive and practical value.

4.2. Comparison of Machine Learning Models for Predicting Quinoa Grain Yield

The results of this study demonstrated clear differences in the performance of machine learning models used for predicting quinoa grain yield. The Multiple Linear Regression model achieved the best performance among all evaluated models, with an R^2 value of 0.9968 and the lowest values of RMSE (0.0176 ton ha^{-1}) and MAE (0.0142 ton ha^{-1}), indicating that the model explained approximately 99.7% of the variation in grain yield with very low prediction errors.

This result may appear unexpected because recent literature often reports superior performance of nonlinear machine learning models, such as Random Forest and XGBoost, in agricultural prediction problems. However, model performance does not depend solely on algorithm complexity; rather, it is strongly influenced by data characteristics, sample size, and the nature of relationships between predictor variables and the target variable (James et al., 2021; Géron, 2022).

In the present study, the dataset consisted of only 18 observations, which represents a relatively small sample size for complex machine learning algorithms. Under such conditions, linear models often provide more stable relationships without requiring the estimation of a large number of parameters, whereas nonlinear models generally require larger and more diverse datasets to reliably learn complex patterns (Hastie et al., 2009).

Furthermore, the results suggest that the relationships between measured agronomic traits and grain yield were largely linear, which explains the exceptional performance of the linear regression model. This interpretation is also supported by the strong correlations observed between grain yield and directly related yield components, particularly grain weight per plant and dry matter yield, allowing the linear model to efficiently represent these relationships without requiring more complex approaches.

These findings are consistent with Shahhosseini et al. (2021), who emphasized that more complex models do not necessarily outperform simpler models and that model selection should depend on data characteristics rather than algorithm complexity. Similarly, van Klompenburg et al. (2020) reported in

their systematic review that machine learning performance in crop yield prediction varies considerably according to dataset size and predictor variables, and that no single model consistently performs best under all conditions.

The results highlight the importance of avoiding the assumption that more advanced algorithms will always provide better predictions. In this study, the simplest model achieved the highest accuracy when applied to a limited agricultural dataset characterized by clear linear relationships. This confirms the necessity of selecting models according to the actual properties of the available data rather than their popularity or mathematical complexity.

4.3. Discussion of Random Forest Model Performance

The Random Forest Regression model ranked third among the evaluated models, achieving an R^2 value of 0.8210, with RMSE and MAE values of 0.1316 and 0.1049 ton ha^{-1}, respectively. These results indicate relatively good predictive ability; however, its performance was lower than that of Multiple Linear Regression and XGBoost.

Random Forest is based on constructing a large number of decision trees and combining their predictions to generate a more stable model. It is characterized by its ability to represent nonlinear relationships and complex interactions among predictor variables (Breiman, 2001). Nevertheless, the efficiency of this model is strongly dependent on dataset size and variability, as sufficient observations are required to construct diverse trees capable of capturing different potential patterns.

In the present study, the limited number of observations may have restricted the ability of Random Forest to fully exploit its advantages. Most decision trees were constructed using training subsets that were highly similar due to the small dataset size, which reduced variability among trees and consequently limited improvements in prediction accuracy compared with the linear model.

Despite this limitation, Random Forest demonstrated good capability in identifying the relative importance of variables affecting grain yield, providing additional insights into the major factors influencing crop productivity. Therefore, the value of this model in the current study is not limited to prediction accuracy but also extends to interpretation of the contribution of different agronomic traits to grain yield, which is particularly valuable in applied studies related to crop breeding and management.

These findings are consistent with the observations of Liakos et al. (2018) and van Klompenburg et al. (2020), who reported that Random Forest performance generally improves with increasing dataset size and number of observations, whereas its efficiency may decline when applied to small datasets compared with simpler models.

4.4. Discussion of XGBoost Model Performance

The XGBoost Regression model ranked second among the evaluated models, achieving an R^2 value of 0.8403, while RMSE and MAE values were 0.1244 and 0.0848 ton ha^{-1}, respectively. These results were superior to those obtained with Random Forest but remained lower than the performance of the linear regression model.

The XGBoost algorithm is based on the Gradient Boosting approach, in which decision trees are constructed sequentially, with each new tree correcting the errors produced by previous trees. This

provides the model with a strong ability to capture nonlinear relationships and complex interactions among variables (Chen & Guestrin, 2016).

Although XGBoost is considered one of the most successful machine learning algorithms for agricultural prediction applications, its performance depends strongly on dataset size and diversity. In small datasets, insufficient information may be available to construct a large number of trees capable of detecting complex patterns, thereby reducing the expected advantage of the algorithm compared with simpler models.

Furthermore, the use of regularization techniques within XGBoost to reduce overfitting may sometimes lead to excessive simplification of the model when dealing with limited datasets. This may explain the lower accuracy of XGBoost compared with linear regression in the present study.

These results agree with Chen and Guestrin (2016) and the review of Shahhosseini et al. (2021), who indicated that XGBoost generally achieves superior performance when applied to large datasets or multi-source datasets, whereas its full potential may not be achieved when using small datasets with predominantly linear relationships.

Although XGBoost was not the best-performing model in this study, it provided relatively high predictive accuracy, suggesting its potential usefulness in future studies involving larger datasets and additional variables, such as climatic, spectral, or genetic information, which may increase the complexity of relationships among predictors.

4.5. Discussion of Support Vector Regression Model Performance

The Support Vector Regression (SVR) model showed the lowest performance among the four evaluated models, with an R^2 value of 0.6784, and RMSE and MAE values of 0.1765 and 0.1399 ton ha^{-1}, respectively. These results indicate a lower ability to represent the relationship between agronomic variables and grain yield compared with the other models.

SVR is known for its ability to model nonlinear relationships using kernel functions and has demonstrated high efficiency in many prediction applications when its parameters are appropriately optimized (Smola & Schölkopf, 2004). However, the performance of this model depends greatly on the selection of key parameters, including the penalty parameter (C), kernel width (γ), and tolerance parameter (ε).

In this study, standard model settings were used to ensure a fair comparison among models without performing extensive hyperparameter optimization. This may have contributed to the lower performance of SVR, particularly considering the limited number of observations.

Moreover, the results indicate that the relationships between input variables and grain yield were closer to linear than highly complex nonlinear relationships. Therefore, the relative advantage provided by SVR through nonlinear kernel functions was limited under the conditions of this study.

These findings are consistent with Liakos et al. (2018), who reported that the success of SVR in agricultural applications depends strongly on data characteristics and parameter optimization procedures. Performance may decrease when small datasets prevent stable estimation of optimal decision boundaries.

Although SVR showed the lowest predictive accuracy in the present study, this does not diminish its importance in other applications. Previous studies have demonstrated its effectiveness when larger datasets are available or when combined with optimization techniques and advanced feature selection approaches. This confirms that algorithm performance is mainly determined by data characteristics rather than by the algorithm itself.

4.6. Why Did the Linear Regression Model Outperform Other Models?

One of the most important findings of this study was that the mathematically simplest model, Multiple Linear Regression, outperformed three widely used advanced machine learning models, namely Random Forest, XGBoost, and Support Vector Regression.

This result highlights an important principle in machine learning: higher algorithmic complexity does not necessarily guarantee improved prediction accuracy. The selection of the optimal model should be based on the characteristics of the dataset rather than on the popularity, novelty, or complexity of the algorithm.

Several interconnected factors may explain this superior performance. The first factor is the small size of the dataset (only 18 observations), as complex machine learning algorithms generally require a large number of observations to effectively learn nonlinear patterns. In contrast, linear regression can establish stable relationships using limited datasets when the relationships among variables are clear.

The second factor is related to the biological nature of the measured traits. Most variables included in this study, such as grain weight per plant, dry matter yield, and harvest index, represent direct or indirect components of final crop yield and are therefore expected to be associated with grain yield through relatively linear relationships. Consequently, the linear regression model was able to represent these relationships with very high accuracy.

The results also suggest that the use of more complex models did not provide additional information beyond that already captured by the linear model. Instead, they may have introduced greater variability in predictions due to the limited amount of available data, explaining their lower performance compared with the linear approach.

This finding carries an important practical message for agricultural researchers: traditional statistical models still retain considerable scientific value, and they may represent the most appropriate choice for studies based on small-scale field experiments. Therefore, model selection should be guided by data characteristics rather than by model complexity.

Overall, this study confirms that integrating agricultural knowledge with appropriate statistical model selection may be more important than applying advanced machine learning algorithms that are not suitable for the available dataset. This conclusion agrees with recent trends in agricultural artificial intelligence research, which emphasize selecting the appropriate algorithm for the specific problem rather than applying complex models indiscriminately (van Klompenburg et al., 2020; Shahhosseini et al., 2021).

4.7. Statistical Comparison Between Models Using the Wilcoxon Signed-Rank Test

The comparison among machine learning models in this study was not limited to conventional performance metrics. The Wilcoxon Signed-Rank Test was also applied to compare the absolute prediction errors of the best-performing model, Multiple Linear Regression, with those of the other evaluated models.

The test results showed highly significant differences ($P < 0.001$) between the prediction errors of the linear regression model and those of Random Forest, Support Vector Regression, and XGBoost. These findings confirm that the superior performance of the linear model was not simply due to minor numerical differences in R^2, RMSE, or MAE values, but rather reflected a consistent and statistically significant improvement in prediction accuracy across all observations.

This result is particularly important because many machine learning studies rely solely on performance indicators without statistically testing whether differences between models are significant. In contrast, the present study provided statistical evidence supporting model selection, thereby increasing the reliability of the conclusions.

This approach agrees with recent recommendations emphasizing the importance of applying appropriate statistical tests when comparing machine learning algorithms, as performance metrics alone may not always be sufficient for determining the superiority of a specific model, particularly when using small datasets or when models show similar performance levels (Demšar, 2006; Benavoli et al., 2017).

These results indicate that Multiple Linear Regression not only achieved the highest explanatory ability and the lowest prediction errors but also represented the most stable and consistent model throughout the cross-validation process. Therefore, it can be considered the most suitable model for predicting quinoa grain yield under the conditions of this study.

4.8. Practical Implications and Future Perspectives

The findings of this study demonstrate that the selection of a machine learning model should be based on dataset characteristics and the nature of the research problem rather than algorithm complexity. Despite the rapid development of artificial intelligence models in recent years, the results showed that Multiple Linear Regression provided the highest predictive performance when applied to a limited field dataset characterized by relatively linear relationships among variables.

The importance of this finding is particularly relevant in agricultural research, where many field experiments rely on a limited number of experimental units due to constraints related to land availability, cost, and time. Under such conditions, traditional statistical models may be more appropriate than some complex machine learning algorithms. This challenges the common assumption that more advanced models always provide superior performance.

From an applied perspective, this study demonstrates that integrating agronomic traits directly associated with yield components with predictive analytical techniques provides an effective tool for supporting crop management decisions. The results can also guide plant breeding programs by focusing on traits that demonstrated the highest importance in explaining variation in grain yield, thereby improving selection efficiency and enhancing quinoa productivity under deficit irrigation conditions.

Despite the promising performance of the evaluated models, future research should focus on developing larger and more diverse datasets including multiple growing seasons, different environmental locations, and additional genotypes. Furthermore, incorporating weather data, soil characteristics, remote sensing-derived spectral indices, and molecular markers is expected to enhance the ability of machine learning models, particularly nonlinear algorithms, to capture complex relationships and improve prediction accuracy under variable agricultural environments.

Overall, this study confirms that successful application of artificial intelligence in agriculture does not depend on using the most complex algorithm, but rather on developing a model that matches the characteristics of the data while achieving a balance between accuracy, interpretability, and practical applicability. This perspective represents an important step toward developing intelligent agricultural systems that support sustainable water management and improve crop production in arid and semi-arid environments.

5. Conclusions

This study demonstrated the effectiveness of integrating machine learning techniques with conventional statistical approaches to improve prediction accuracy and enhance the interpretation of agronomic performance traits. The evaluated models provided valuable insights into the relationships between environmental factors, management practices, and crop traits, including relationships that may be difficult to characterize using traditional statistical methods alone.

The application of feature selection techniques contributed to identifying the most influential factors affecting prediction performance, thereby improving model interpretability and reducing unnecessary complexity without compromising predictive ability. The findings confirm that artificial intelligence techniques should not be considered alternatives to conventional statistical analysis, but rather complementary tools that provide advanced analytical perspectives for understanding complex agricultural datasets.

Overall, this study highlights that machine learning methodologies represent a promising approach for accelerating data-driven decision-making in agricultural research, particularly in crop breeding programs, evaluation of crop performance under variable environmental conditions, and development of climate-smart agricultural strategies.

Practical and Research Recommendations

First: Practical Recommendations

- Machine learning models are recommended as supportive tools for agricultural decision-making, particularly in crop variety evaluation programs and the identification of superior genetic materials under different environmental conditions.
- The integration of artificial intelligence models with conventional field experiments can improve data analysis efficiency and reduce the time and effort required to evaluate differences among treatments or genotypes.
- Feature importance analysis should be considered as a valuable approach for identifying the most influential factors affecting productivity and agronomic traits, thereby supporting better decisions related to crop management and resource optimization.

- The proposed analytical framework can serve as a foundation for developing intelligent decision-support systems in precision agriculture, particularly in regions exposed to climatic variability and challenging environmental conditions.

Second: Research Recommendations

- Future studies should expand the datasets used for machine learning modeling by including multiple growing seasons, diverse geographical locations, and a larger number of varieties or genetic materials to enhance model generalization capability.
- Comparative studies involving different machine learning algorithms and deep learning approaches are recommended to determine the most suitable models for different types of agricultural datasets.
- Additional data sources should be integrated, including remote sensing information, soil properties, detailed climatic variables, and genetic information, to develop more comprehensive predictive models.
- Greater attention should be given to Explainable Artificial Intelligence (XAI) techniques, such as SHAP (SHapley Additive exPlanations) and Partial Dependence Analysis, to establish links between machine learning predictions and the biological interpretation of agricultural processes.

Strengths and Limitations of the Study

Strengths

This study has several important strengths, most notably the development of an integrated analytical framework combining the advantages of conventional statistical analysis with the advanced predictive capabilities of machine learning algorithms. This integration provided deeper insights into relationships among variables and improved the ability to predict the target trait.

Furthermore, the use of feature selection approaches enhanced model efficiency and interpretability by identifying the variables that contributed most to prediction performance. In addition, the methodology developed in this study provides a framework that can be applied and expanded to other crops and agricultural environments, increasing its practical value in the fields of smart and sustainable agriculture.

Limitations

Despite the promising results obtained, several limitations should be considered. The performance of machine learning models is highly dependent on the quantity, quality, and diversity of available data. Therefore, increasing the number of experiments, locations, and growing seasons would improve model reliability and generalization capability.

Moreover, some advanced machine learning models, despite their high predictive capacity, may provide less transparent interpretations of relationships among variables compared with traditional statistical models. Therefore, integrating these models with Explainable Artificial Intelligence approaches represents an important future research direction.

Finally, the proposed models require further validation using independent datasets collected from different geographical environments before their widespread adoption in agricultural decision-support systems.

References

Awa, M., Zhao, J., & Tumaerbai, H. (2024). Deficit irrigation-based improvement in growth and yield of quinoa in the northwestern arid region in China. *Sustainability, 16*(10), 4136. https://doi.org/10.3390/su16104136

Bazile, D., Jacobsen, S.-E., & Verniau, A. (Eds.). (2016). *State of the Art Report on Quinoa around the World in 2013*. Food and Agriculture Organization of the United Nations (FAO).

Benavoli, A., Corani, G., Demšar, J., & Zaffalon, M. (2017). Time for a change: A tutorial for comparing multiple classifiers through Bayesian analysis. *Journal of Machine Learning Research, 18*(77), 1–36.

Dehghanian, Z., Ahmadabadi, M., Asgari Lajayer, B., Gougerdchi, V., Hamedpour-Darabi, M., Bagheri, N., Sharma, R., Vetukuri, R. R., Astatkie, T., & Dell, B. (2024). Quinoa: A promising crop for resolving the bottleneck of cultivation in soils affected by multiple environmental abiotic stresses. *Plants, 13*(15), 2117. https://doi.org/10.3390/plants13152117

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*, 1–30.

Geerts, S., & Raes, D. (2009). Deficit irrigation as an on-farm strategy to maximize crop water productivity in dry areas. *Agricultural Water Management, 96*(9), 1275–1284. https://doi.org/10.1016/j.agwat.2009.04.009

Geerts, S., Raes, D., Garcia, M., Taboada, C., Miranda, R., Cusicanqui, J., Mhizha, T., & Vacher, J. (2009). Modeling the potential for closing quinoa yield gaps under varying water availability in the Bolivian Altiplano. *Agricultural Water Management, 96*(11), 1652–1658.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (3rd ed.). O'Reilly Media.

Govaichelvan, K. N., Pathmanathan, D., Zainal-Abidin, R.-A., & Abu, A. (2024). Machine learning for major food crops breeding: Applications, challenges, and ways forward. *Agronomy Journal, 116*(3), 1112–1125. https://doi.org/10.1002/agj2.21393

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Hirich, A., Choukr-Allah, R., & Jacobsen, S.-E. (2014a). Deficit irrigation and organic compost improve growth and yield of quinoa and pea. *Journal of Agronomy and Crop Science, 200*(5), 390–398. https://doi.org/10.1111/jac.12073

Hirich, A., Choukr-Allah, R., & Jacobsen, S.-E. (2014b). The combined effect of deficit irrigation by treated wastewater and organic amendment on quinoa (*Chenopodium quinoa* Willd.) productivity. *Desalination and Water Treatment, 52*(10–12), 2208–2213. https://doi.org/10.1080/19443994.2013.777944

Hirich, A., Jelloul, A., Choukr-Allah, R., & Jacobsen, S.-E. (2014c). Saline water irrigation of quinoa and chickpea: Seedling rate, stomatal conductance and yield responses. *Journal of Agronomy and Crop Science, 200*(5), 378–389. https://doi.org/10.1111/jac.12072

Jabed, M. A., & Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon, 10*(24), Article e40836. https://doi.org/10.1016/j.heliyon.2024.e40836

Jacobsen, S. E., Mujica, A., & Jensen, C. R. (2003). The resistance of quinoa (*Chenopodium quinoa* Willd.) to adverse abiotic factors. *Food Reviews International, 19*(1–2), 99–109.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer.

Khatibi, S. M. H., & Ali, J. (2024). Harnessing the power of machine learning for crop improvement and sustainable production. *Frontiers in Plant Science, 15*, Article 1417912. https://doi.org/10.3389/fpls.2024.1417912

Kuhn, M., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674. https://doi.org/10.3390/s18082674

Mirsafi, S. M., Sepaskhah, A. R., & Ahmadi, S. H. (2024a). Physiological traits, crop growth, and grain quality of quinoa in response to deficit irrigation and planting methods. *BMC Plant Biology*, *24*, 809. https://doi.org/10.1186/s12870-024-05523-5

Mirsafi, S. M., Sepaskhah, A. R., & Ahmadi, S. H. (2024b). Quinoa growth and yield, soil water dynamics, root growth, and water use indicators in response to deficit irrigation and planting methods. *Journal of Agriculture and Food Research, 15*, 100970. https://doi.org/10.1016/j.jafr.2024.100970

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis* (6th ed.). Wiley.

Ruiz, K. B., Biondi, S., Oses, R., Acuña-Rodríguez, I. S., Antognoni, F., Martinez-Mosqueira, E. A., et al. (2016). Quinoa biodiversity and sustainability for food security under climate change. *Agronomy*, *6*(1), 1–18.

Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction: A review. *Agronomy, 11*, 1619.

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture, 177*, 105709.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*(1), 79–82.