

Article

A Simplified Evaluation Model for Soybean Seedling Salt Tolerance Based on Core Biomass Traits Under Saline Pond Conditions

Yixin Tian, Jinying Zhu *, Fangjing Hua, Pengpeng Cao, Chunyan Li, Chunyu Wang, Guanxiong Zhu, Qi Gao and Fengju Gao

Dezhou Academy of Agricultural Sciences, Shandong Research Institute of Rural Revitalization Practice, Dezhou 253015, China; dznkyzlstyx@dz.shandong.cn (Y.T.); dznkyzslshfj@dz.shandong.cn (F.H.); dznkyjdk@dz.shandong.cn (P.C.); dznkyzslscy@dz.shandong.cn (C.L.); dznkyzslswcy@dz.shandong.cn (C.W.); dznkyzslszgx@dz.shandong.cn (G.Z.); dznkyzslsgq@dz.shandong.cn (Q.G.); dznkygfj@dz.shandong.cn (F.G.)

* Correspondence: dznkyzls@dz.shandong.cn; Tel.: +86-0534-2321660

Abstract

Salt stress severely restricts soybean seedling growth and yield formation, and the redundant indicators and low efficiency of conventional salt tolerance evaluation methods limit the large-scale screening and breeding of salt-tolerant soybean germplasms. In this study, we aimed to establish a simplified and efficient salt tolerance evaluation system for soybean seedlings under saline pond conditions. We determined 16 phenotypic traits of 100 soybean germplasm resources under soil salt stress (0.3% soil salt content, EC 5.0 dS/m), calculated the salt tolerance coefficient (STC) of each trait, and comprehensively analyzed phenotypic variation, correlation, germplasm classification, and core evaluation indices via principal component analysis (PCA), K-means clustering, random forest model, SHAP interpretation, 10-fold nested cross-validation, and correlation network analysis. Biomass-related traits exhibited abundant phenotypic variation, with coefficients of variation ranging from 42.6% to 48.4%. Five principal components explained 82.90% of the total phenotypic variation and divided the accessions into four salt tolerance categories. Total fresh weight (TFW), stem fresh weight (SFW), and leaf fresh weight (LFW) were identified as the core indices, together accounting for over 93% of the total feature importance in the random forest model, whereas the remaining 13 traits each contributed less than 1.2%. The simplified three-index model showed strong consistency with the full 16-trait model (Pearson $r > 0.970$, AUC = 0.970) and achieved a screening accuracy of 90.0% under 10-fold nested cross-validation. Under the experimental conditions examined, fresh biomass accumulation emerged as the dominant phenotypic characteristic associated with seedling salt tolerance. This simplified evaluation framework may facilitate rapid preliminary screening of salt-tolerant soybean germplasms at the seedling stage, pending further validation across diverse environments and genetic backgrounds.



Academic Editor: Junhua Peng

Received: 22 May 2026

Revised: 19 July 2026

Accepted: 19 July 2026

Published: 22 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: soybean; seedling stage; salt stress; phenotypic trait; biomass; random forest; SHAP analysis; salt tolerance evaluation

1. Introduction

Soil salinity is a major abiotic stress limiting soybean (*Glycine max* L.) growth, development, and yield. Approximately 20% of irrigated farmland worldwide is affected by salinization, with this proportion increasing annually [1,2]. As an important oil and protein

crop, soybean is relatively sensitive to salt stress. Salt damage at the seedling stage significantly reduces biomass and ultimately impacts yield [3,4]. Thus, systematically analyzing phenotypic variation in soybean salt tolerance, identifying core evaluation indicators, and establishing an efficient evaluation system are critical for screening salt-tolerant germplasm and improving breeding efficiency.

Soybean salt tolerance is a complex quantitative trait controlled by multiple genes, making direct evaluation challenging. Previous studies have typically used multiple indicators (e.g., survival rate, plant height, biomass, SPAD (Soil Plant Analysis Development, relative chlorophyll content), ion content) for comprehensive evaluation [5–8]. However, due to varying correlations and information redundancy among these indicators, full-indicator evaluation systems are cumbersome, costly, and unsuitable for large-scale germplasm screening. In recent years, principal component analysis (PCA), cluster analysis, and machine learning have been widely used in crop stress tolerance evaluation [9–11], providing new technical approaches for indicator dimensionality reduction and core trait screening. Among these methods, Random Forest effectively handles high-dimensional data and quantifies the relative contribution of each trait to model predictions [12,13], while SHAP (SHapley Additive exPlanations) analysis clarifies the direction and strength of associations between features and model outputs [14,15]. Their combination supports core salt tolerance indicator mining and model interpretation. While this approach has been successfully applied in maize drought tolerance and sorghum salt tolerance studies [16,17], systematic application to identify core salt tolerance indicators in soybean remains limited.

Current soybean salt tolerance evaluation research faces three key issues. First, evaluation indicators vary widely across studies, lacking a unified core system [6,18,19]. Second, the relative contributions of different indicator dimensions (e.g., biomass, water status, photosynthesis) to predicting salt tolerance scores remain unclear [20]. Third, most existing evaluation systems rely on full-indicator comprehensive scores, which are complex to operate and difficult to promote in breeding practice [21]. Identifying the primary predictors of soybean salt tolerance and establishing a simplified, efficient evaluation system are therefore urgent tasks in current soybean salt-tolerant breeding.

Based on this background, this study used 100 soybean germplasm accessions and measured 16 salt tolerance-related phenotypic traits at the seedling stage under saline pond conditions (0.3% soil salt content, EC 5.0 dS/m). The objectives were to (1) analyze the phenotypic variation and correlation structure of salt tolerance traits; (2) construct a comprehensive soybean salt tolerance evaluation model using PCA; (3) identify core evaluation indicators via Random Forest and SHAP analysis; (4) establish and internally validate a simplified evaluation model based on core indicators; and (5) provide a methodological framework that integrates PCA-based germplasm classification with machine learning-based trait weight calibration for identifying core salt tolerance indicators at the seedling stage. In this framework, the PCA-derived composite score serves as a standardized metric for salt tolerance grading, while Random Forest and SHAP analyses complement the linear PCA weighting by quantifying nonlinear trait contributions, thereby enabling objective screening of core evaluation indicators.

2. Materials and Methods

2.1. Plant Materials and Experimental Site

A total of 100 soybean (*Glycine max* L.) germplasm accessions were used as experimental materials, sourced from the Germplasm Resource Platform of the National Soybean Industry Technology System. Detailed information on these accessions, including maturity type, geographic origin, and ecoregion, is presented in Supplementary Table S1. All

materials were evaluated at the experimental station of Dezhou Academy of Agricultural Sciences (116°21' E, 37°27' N) from June to October 2025.

The experimental site is located in a warm-temperate semi-humid monsoon climate zone, with a mean annual temperature of 13.1 °C and mean annual precipitation of 556 mm, concentrated primarily from June to August. Monthly mean temperature and precipitation during the experimental period are provided in Supplementary Table S2. The soil is classified as loamy fluvo-aquic soil. Basic physicochemical properties of the 0–20 cm soil layer before the experiment are presented in Table 1.

Table 1. Soil physicochemical properties of the experimental site (0–20 cm layer) before sowing.

Parameter	Salt-Stressed Pond	Control Pond
Soil texture	Loam	Loam
Organic matter (g/kg)	10.5 ± 0.8	10.8 ± 0.7
Total N (g/kg)	0.65 ± 0.05	0.68 ± 0.04
Available P (mg/kg)	18.2 ± 1.5	17.8 ± 1.3
Available K (mg/kg)	185 ± 12	180 ± 10
pH	8.2 ± 0.2	7.5 ± 0.1
EC (dS/m)	5.0 ± 0.3	0.6 ± 0.1
Total salt content (%)	0.30 ± 0.02	<0.05

The trial was conducted in artificially constructed ponds. To establish the salt-stressed environment, salt was evenly spread on the soil surface and incorporated into the 0–20 cm soil layer via tillage before sowing. This procedure adjusted the total soil salt content to 0.3% (EC 5.0 dS/m). Throughout the experimental period, soil salt content and EC were monitored every 15 days at five fixed sampling points per pond (four corners and center) at two soil depths (0–10 cm and 10–20 cm). When soil moisture fell below 60% of field capacity, drip irrigation using groundwater was applied uniformly across all plots. Soil EC was re-measured 48 h after each irrigation event to verify salinity stability.

To assess spatial uniformity within each treatment, coefficients of variation (CV) for soil salt content and EC were calculated across the five sampling points at each monitoring time. The CV of total salt content ranged from 5.2% to 8.7%, and the CV of EC ranged from 4.8% to 7.5%, both below 10%, confirming satisfactory spatial homogeneity within the salt-stressed ponds. No statistically significant differences in EC or pH were detected among the three blocks ($p > 0.05$), further supporting the environmental uniformity across replicates.

Control plots were arranged alongside salt-stressed plots within the same randomized complete block design. No additional salt was applied to the control plots, where the total soil salt content was below 0.05% (pH 7.5, EC 0.6 dS/m). These low-salinity conditions were considered unlikely to inhibit soybean growth.

Each pond measured 6.5 m in length and 5 m in width. Two treatments—0.3% soil salt stress (EC 5.0 dS/m) and the untreated control—were arranged in a randomized complete block design with three blocks. Each block contained both treatments, with one plot per treatment per block. The three blocks served as three biological replicates. Each plot covered an area of 16 m² and contained 16 sown rows of soybean, with a row length of 2.5 m, a row spacing of 0.4 m, and a plant spacing of 0.1 m. The experimental plot served as the experimental unit. Uniform field management practices were adopted across all plots to ensure consistent conditions regarding sunlight, water supply, tillage, and plant protection.

2.2. Trait Measurement

Traits were measured when each accession reached the V1 stage (first trifoliate leaf fully expanded). Because emergence time varied among accessions, sampling was performed by developmental stage rather than on a fixed day after sowing. Across accessions within each

treatment, the sampling period spanned no more than three days, ensuring comparable developmental staging for all entries.

Three uniform seedlings per plot were selected for destructive sampling. The mean value of these three plants was calculated and used as the plot-level observation. With 100 accessions, 2 treatments, and 3 blocks, a total of 600 plot-level observations were obtained for each trait, derived from 1800 individual plant measurements.

2.2.1. Growth Traits

Survival rate (SR, %): Calculated as the percentage of surviving plants relative to the total number of sown plants per plot at the time of trait measurement.

Root length (RL, cm): Measured from the root crown to the apical meristem using a ruler (precision 0.1 cm).

Plant height (PH, cm): Measured from the soil surface to the apical meristem using a ruler (precision 0.1 cm).

2.2.2. Photosynthetic Trait

Relative chlorophyll content (SPAD value): Measured using a SPAD-502Plus chlorophyll meter (Konica Minolta, Osaka, Japan). Three measurements were taken at the middle portion of the fully expanded first trifoliate leaf, and the mean value was recorded.

2.2.3. Biomass Traits

Plants were separated into roots, stems, and leaves. Root samples were excavated to a depth of 30 cm using a standardized soil excavation protocol. The excavated root systems were gently washed with running water over a 0.5 mm mesh sieve to remove adhering soil while minimizing root loss. Excess water was removed with absorbent paper before weighing.

Fresh weights—root fresh weight (RFW, g), stem fresh weight (SFW, g), and leaf fresh weight (LFW, g)—were determined using an electronic analytical balance (ME204E, Mettler Toledo, Greifensee, Switzerland; precision: 0.1 mg). Total fresh weight (TFW, g) was calculated as the sum of RFW, SFW, and LFW.

Dry weights were measured after oven-drying at 105 °C for 30 min for enzyme deactivation, followed by drying at 80 °C to constant weight in a forced-air drying oven (DHG-9240A, Yiheng, Shanghai, China). Root dry weight (RDW, g), stem dry weight (SDW, g), and leaf dry weight (LDW, g) were recorded, and total dry weight (TDW, g) was calculated as their sum.

Root-to-shoot ratio (RSR) was calculated as $RDW / (SDW + LDW)$.

2.2.4. Water-Related Traits

The following water-related traits were calculated using the respective formulas:

Root water content (RWC, %) = $(RFW - RDW) / RFW \times 100\%$.

Shoot water content (SWC, %) = $[(SFW + LFW - SDW - LDW) / (SFW + LFW)] \times 100\%$.

Fresh-to-dry weight ratio (FWR) = TFW / TDW .

2.3. Data Statistical Analysis

Data organization and statistical analyses were performed using Microsoft Excel 2021 (Microsoft Corp., Redmond, WA, USA), SPSS 26.0 (IBM Corp., Armonk, NY, USA), R 4.2.0 (R Foundation for Statistical Computing, Vienna, Austria), and Python 3.9 (Python Software Foundation, Wilmington, DE, USA).

The 16 evaluated traits included both directly measured traits and derived indices (e.g., water content, root-to-shoot ratio, fresh-to-dry weight ratio) that capture additional salt tolerance characteristics. To address the inherent correlations and data redundancy among these traits, a sequential analytical strategy was adopted: correlation analysis was first applied to quantify redundancy, followed by PCA and multivariate methods for dimensionality reduction and key indicator screening.

As detailed in Section 2.3.1, the salt tolerance coefficient (STC) was calculated for each trait with standard error estimated from three biological replicates. Two-way ANOVA was performed to test the effects of accession, treatment, block, and their interaction. Data normality and homogeneity of variances were assessed using descriptive measures (skewness, kurtosis, and kernel density curves) and Levene's test, respectively. A statistically significant accession $\times$ treatment interaction ($p < 0.05$) confirmed accession-specific salt tolerance responses, justifying subsequent STC-based evaluation. Differences between treatment groups were assessed using one-way ANOVA followed by Tukey's HSD test ($p < 0.05$).

All subsequent multivariate analyses—including correlation analysis, PCA, K-means clustering, and random forest modeling—were conducted on mean STC values. Detailed methodologies are provided in the respective sections below.

2.3.1. Calculation of Salt Tolerance Coefficient

To eliminate differences in basal growth vigor among germplasm accessions, the salt tolerance coefficient (STC) for each trait was calculated as

$$\text{STC} = \text{Measured value under salt stress treatment} / \text{Measured value under control treatment}$$

For all traits, STC = 1.0 indicates no salt effect, STC < 1.0 indicates salt-induced reduction, and STC > 1.0 indicates salt-induced enhancement. For growth-related traits, higher STC values generally reflect greater salt tolerance.

In this randomized complete block design, STC was calculated based on paired data within each block. For each accession, the STC of one block was obtained by dividing the phenotypic value of the salt-stressed plot by the value of the paired control plot in the identical block. Three STC values corresponding to three blocks (three biological replicates) were generated for each accession. The mean STC across the three blocks was then calculated and used as the salt tolerance index for each accession–trait combination. The standard error (SE) of the STC for each accession was estimated as

$$SE = SD / \sqrt{3}$$

where SD is the standard deviation of STC values across the three biological replicates. This approach preserves replicate-level information and provides an estimate of the uncertainty associated with each STC measurement. The mean STC was used for all subsequent statistical analyses.

A two-way analysis of variance (ANOVA) was performed to confirm that different accessions exhibited differential responses to salt stress—an essential prerequisite for evaluating accession-specific salt tolerance—using the following linear model:

$$Y = \mu + G_i + T_j + B_k + (G \times T)_{ij} + \varepsilon_{ijkl}$$

where Y is the observed trait value of the i -th accession under the j -th treatment in the k -th block; μ is the overall mean; G_i is the fixed effect of the i -th accession; T_j is the fixed effect of the j -th treatment (salt stress or control); B_k is the random effect of the k -th block; $(G \times T)_{ij}$ is the fixed interaction effect between accession and treatment; and ε_{ijkl} is

the random residual error. A significant accession $\times$ treatment interaction ($p < 0.05$) was considered evidence of accession-specific salt tolerance responses.

2.3.2. Phenotypic Variation and Correlation Analysis

Descriptive statistics of STC values for the 16 traits were calculated, including minimum, maximum, mean, standard deviation, coefficient of variation (CV), skewness, and kurtosis. Pearson correlation coefficients were computed to evaluate pairwise correlations among STC values, and correlation heatmaps were generated using the pheatmap package (version 1.0.12) in R 4.2.0. This analysis aimed to characterize the overall correlation structure and assess information redundancy across traits.

2.3.3. Principal Component Analysis and Comprehensive Score

PCA was performed on the STC values of the 16 traits after Z-score standardization [9,10]. Principal components with eigenvalues > 1 were extracted, and variance contribution rates and cumulative variance contribution rates were calculated. The comprehensive score for each accession was computed as: Comprehensive Score = $\sum(\text{PC}_i \times \text{V}_i) / \sum \text{V}_i$, where PC_i is the score of the i -th principal component and V_i is its variance contribution rate. PCA was implemented using the prcomp function in R (version 4.2.0).

2.3.4. Cluster Analysis

The optimal number of clusters was determined by combining the elbow method and silhouette coefficient analysis. K-means clustering was performed on the standardized scores of the five principal components derived from the 16-trait PCA, based on the optimal K value [11], using the factoextra package (version 1.0.7) in R (version 4.2.0). In this procedure, each principal component was implicitly weighted by its variance contribution. Cluster validity was further assessed using the Calinski–Harabasz index, Davies–Bouldin index, and bootstrap resampling stability analysis.

2.3.5. Random Forest and SHAP Analysis

To complement the inherent limitation of linear PCA weighting, which assumes each trait contributes independently and proportionally to the composite score, Random Forest and SHAP analyses were employed to quantify the nonlinear contributions of individual traits. In this framework, the PCA-derived comprehensive score served as the response variable—not as an independent external outcome to be predicted, but as a standardized internal metric representing the integrated salt tolerance level of each accession. The objective of this machine learning analysis was therefore to calibrate trait weights within the established PCA evaluation framework.

A Random Forest regression model was constructed with the comprehensive score as the dependent variable and the STC values of the 16 traits as independent variables ($n_{\text{estimators}} = 200$, $\text{max_features} = \text{'sqrt'}$, $\text{random_state} = 42$) [12,13]. The model was interpreted using the SHAP method [14,15] with a TreeExplainer. SHAP values were calculated to quantify the marginal contribution of each trait to the comprehensive score, with positive values indicating positive associations and negative values indicating negative associations. Random Forest and SHAP analyses were implemented using the sklearn.ensemble (scikit-learn version 1.0.2) and shap (version 0.41.0) libraries in Python 3.9, respectively.

2.3.6. Simplified Model Construction and Validation

Definition of Two Evaluation Models

Two evaluation systems were established: the full-index model and the simplified model. The full-index model served as the benchmark, constructed from all 16 measured traits. The comprehensive salt tolerance score derived from this model was adopted as

the unified standard for germplasm salt tolerance evaluation and ranking. The simplified model was developed by retaining only the most representative core traits that dominated salt tolerance variation, enabling efficient field identification.

Indicator Screening Procedure for the Simplified Model

Indicator screening was performed hierarchically: traits with high variance contribution to principal components were retained; SHAP values were referenced to select traits with high and stable importance; indicators with strong pairwise correlations were evaluated to minimize residual redundancy; and the biological functions of each trait under salt stress were considered when finalizing the simplified indicator system. Although the three core fresh biomass traits (TFW, SFW, LFW) exhibited strong intercorrelations, they were jointly retained because PCA identified fresh biomass accumulation as the dominant dimension of salt tolerance variation, and the simplified model combined them into a single principal component rather than treating them as independent predictors.

Multi-Indicator Joint Judgment Rule

The optimal threshold for each core indicator was determined by maximizing the Youden index ($J = \text{sensitivity} + \text{specificity} - 1$). An accession was classified as highly salt-tolerant when at least two of the three indicators met or exceeded their respective optimal thresholds. Accessions meeting fewer than two threshold criteria were classified as non-highly salt-tolerant.

Model Validation and Comparison Scheme

Multiple validation approaches were employed. Correlation consistency between the full-index and simplified models was evaluated using Pearson, Spearman, and rank correlation coefficients. ROC curve analysis was performed to calculate AUC, sensitivity, and specificity, evaluating the discriminative ability of the simplified model. To evaluate the stability and generalizability of the simplified model, ten-fold nested cross-validation was adopted [22]: the outer 10-fold loop provided unbiased performance estimates on held-out test subsets, while the inner 10-fold loop optimized PCA component number (1, 2, or 3) and standardization strategy (with or without Z-score) exclusively on training data. A fixed global random seed (SEED = 42) was applied throughout to ensure reproducibility. Feature importance stability across folds was assessed using Random Forest ($n_{\text{estimators}} = 200$, $\text{random_state} = \text{outer_fold}$, $n_{\text{jobs}} = 1$) trained on each outer training set.

2.3.7. Correlation Network Analysis

A correlation network was constructed based on the Pearson correlation coefficient matrix ($|r| > 0.5$, $p < 0.05$) for the key phenotypic traits. The network was visualized using the Fruchterman–Reingold layout in the networkx library (version 2.8) in Python 3.9.

All analytical codes are available from the corresponding author on reasonable request.

3. Results

3.1. Phenotypic Variation Characteristics of Salt Tolerance-Related Traits in Soybean

Under salt stress, the 16 phenotypic traits displayed substantial variation in salt tolerance coefficients (STC), with distinct differences in range, central tendency, dispersion, and distribution shape (Table 2). Core growth traits (SR, RL, PH, SPAD) exhibited moderate STC values (means: 0.58–0.83) and relatively low coefficients of variation (CV: 12.2–25.6%), with most values concentrated in the 0.3–1.0 interval. SPAD and PH showed particularly narrow distributions, indicating limited phenotypic differentiation among accessions for these traits.

Table 2. Descriptive statistics of salt tolerance coefficients (STC) for 16 salt tolerance-related traits in soybean at the seedling stage.

Trait (Abbreviation)	Min	Max	Mean	SD	CV (%)	Skewness	Kurtosis
SR	0.300	0.840	0.578	0.148	25.57	−0.014	−1.116
RL	0.330	1.000	0.657	0.164	24.95	−0.066	−0.934
PH	0.460	0.980	0.767	0.131	17.09	−0.418	−0.760
SPAD	0.550	0.990	0.833	0.102	12.21	−0.792	0.042
RFW	0.130	0.900	0.491	0.210	42.64	0.309	−0.858
SFW	0.110	0.940	0.426	0.200	46.90	0.652	−0.299
LFW	0.110	0.960	0.424	0.205	48.39	0.801	−0.037
RDW	0.130	0.920	0.443	0.202	45.51	0.675	−0.297
SDW	0.110	0.930	0.435	0.209	48.04	0.672	−0.389
LDW	0.070	0.970	0.461	0.220	47.77	0.600	−0.402
TFW	0.120	0.870	0.425	0.186	43.70	0.538	−0.675
TDW	0.090	0.900	0.447	0.199	44.57	0.534	−0.488
RSR	0.420	2.200	1.068	0.405	37.91	0.647	−0.090
RWC	0.480	1.880	1.087	0.258	23.71	0.766	0.993
SWC	0.550	1.330	0.974	0.133	13.60	−0.553	1.824
FWR	0.520	1.790	0.990	0.248	25.02	0.597	0.492

Notes: STC was calculated as the value under salt stress divided by the value under control conditions. SD: standard deviation; CV: coefficient of variation. Trait abbreviations: SR, survival rate; RL, root length; PH, plant height; SPAD, relative chlorophyll content; RFW, root fresh weight; SFW, stem fresh weight; LFW, leaf fresh weight; RDW, root dry weight; SDW, stem dry weight; LDW, leaf dry weight; TFW, total fresh weight; TDW, total dry weight; RSR, root-to-shoot ratio; RWC, root water content; SWC, shoot water content; FWR, fresh-to-dry weight ratio.

Biomass-related traits exhibited the greatest phenotypic variation. STC means for fresh and dry weight traits ranged from 0.43 to 0.49, with CVs of 42.6–48.4% and an overall range spanning 0.07–0.97. LFW and SDW showed particularly pronounced dispersion, reflecting substantial phenotypic differences in biomass accumulation capacity among accessions under salt stress.

Functional traits (RSR, water-related indicators, FWR) displayed intermediate variation patterns, with STC means ranging from 0.97 to 1.09 (CV: 13.6–37.9%) and a broad distribution range (0.42–2.20) that included numerous extreme accessions (STC > 1.5; Figure 1A). Kernel density curves showed that STC values were approximately symmetrically distributed around their central tendencies (Figure 1B). Combined with the skewness and kurtosis values in Table 2 (absolute skewness < 1 for 15 of 16 traits), the data distributions were deemed acceptable for subsequent parametric analyses.

The wide STC range, high inter-accession dispersion, and abundance of extreme phenotypes collectively indicate rich phenotypic variation among accessions within the germplasm panel, providing a robust foundation for salt-tolerant germplasm screening, key indicator identification, and breeding applications.

Two-way ANOVA confirmed statistically significant accession × treatment interactions ($p < 0.01$) for all 16 traits (Supplementary Table S3), demonstrating that the observed STC variation reflects genuine accession-specific responses to salt stress rather than random environmental variation. This significant interaction provided the statistical justification for subsequent comprehensive salt tolerance evaluation and simplified model development.

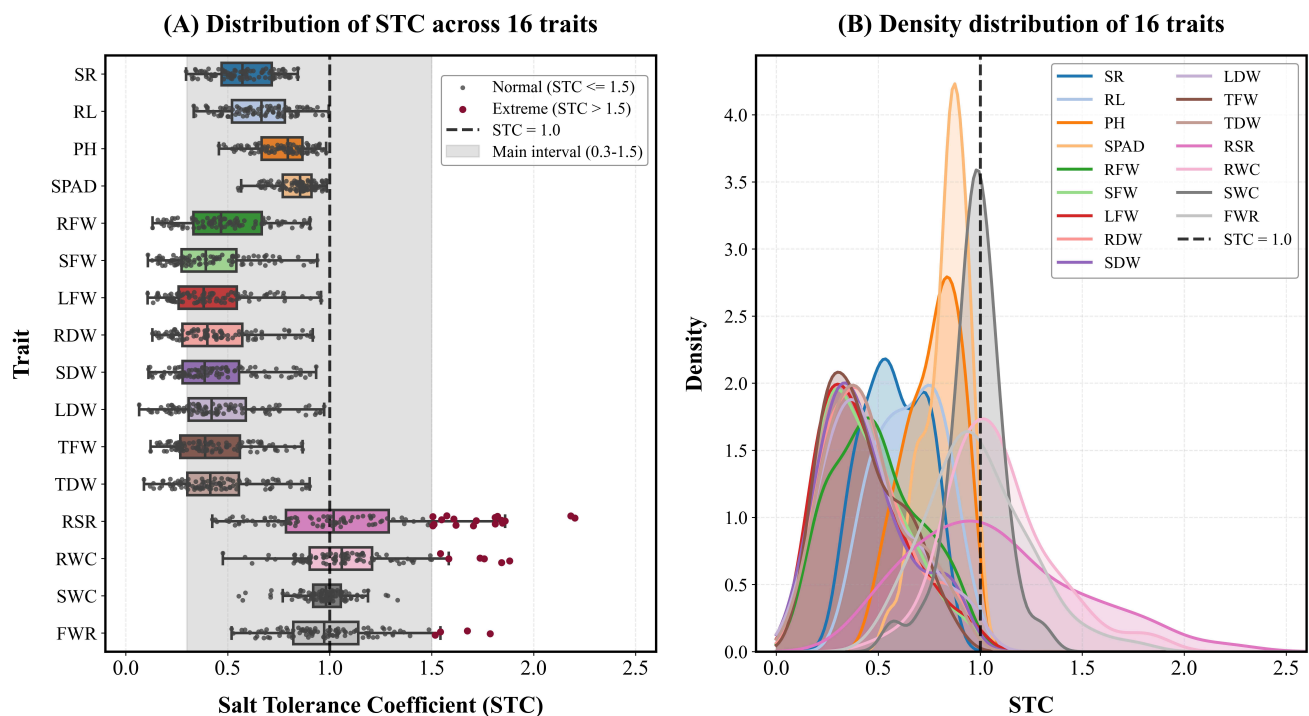


Figure 1. Distribution patterns of salt tolerance coefficients (STC) for 16 seedling traits in soybean. (A) Boxplots of STC values for 16 traits. Boxes represent the median, interquartile range (IQR), and whiskers ($1.5 \times \text{IQR}$). Gray points indicate observations with $\text{STC} \leq 1.5$, and maroon points represent $\text{STC} > 1.5$, used here for visual highlighting of the upper tail of the distribution rather than as a formal statistical threshold for defining extreme salt tolerance. The dashed black line marks $\text{STC} = 1.0$ (no salt effect), and the gray shaded area denotes the main distribution interval (0.3–1.5). (B) Kernel density curves of STC for all 16 traits. Each colored curve represents the distribution of one trait, with peak position and width reflecting the stability and phenotypic variation in salt responses. Traits with left-shifted, broad distributions (e.g., biomass indicators) show stronger salt-induced inhibition and greater phenotypic variation, while traits with peaks near 1.0 and narrow distributions (e.g., SPAD) are more stable under salt stress.

3.2. Correlation Analysis and Information Redundancy Among Salt Tolerance Traits

Pearson correlation analysis of the 16 indicators revealed extensive statistically significant correlations, with pronounced information redundancy among biomass-related traits (Figure 2). TFW was significantly positively correlated with SFW ($r = 0.95$, $p < 0.001$), LFW ($r = 0.98$, $p < 0.001$), and TDW ($r = 0.84$, $p < 0.001$). TDW was similarly strongly correlated with SDW ($r = 0.93$, $p < 0.001$) and LDW ($r = 0.99$, $p < 0.001$). Fresh weight and corresponding dry weight of each organ showed strong positive correlations (root: $r = 0.66$; stem: $r = 0.85$; leaf: $r = 0.79$; all $p < 0.001$), indicating substantial information overlap between these two measurement dimensions.

Morphological indicators displayed weaker associations with biomass traits. PH was moderately correlated with SPAD ($r = 0.35$, $p < 0.001$), and RL showed a moderate correlation with RDW ($r = 0.38$, $p < 0.001$). However, correlations between morphological and biomass-related indicators were generally low ($r = 0.15$ – 0.29), suggesting limited information redundancy across these trait categories. Among water-related indicators, SWC was strongly positively correlated with FWR ($r = 0.92$, $p < 0.001$). RWC showed a weak negative correlation with RDW ($r = -0.35$, $p < 0.001$) and a moderate positive correlation with SWC ($r = 0.25$, $p < 0.05$), indicating limited rather than complete independence of water-related traits from biomass dimensions.

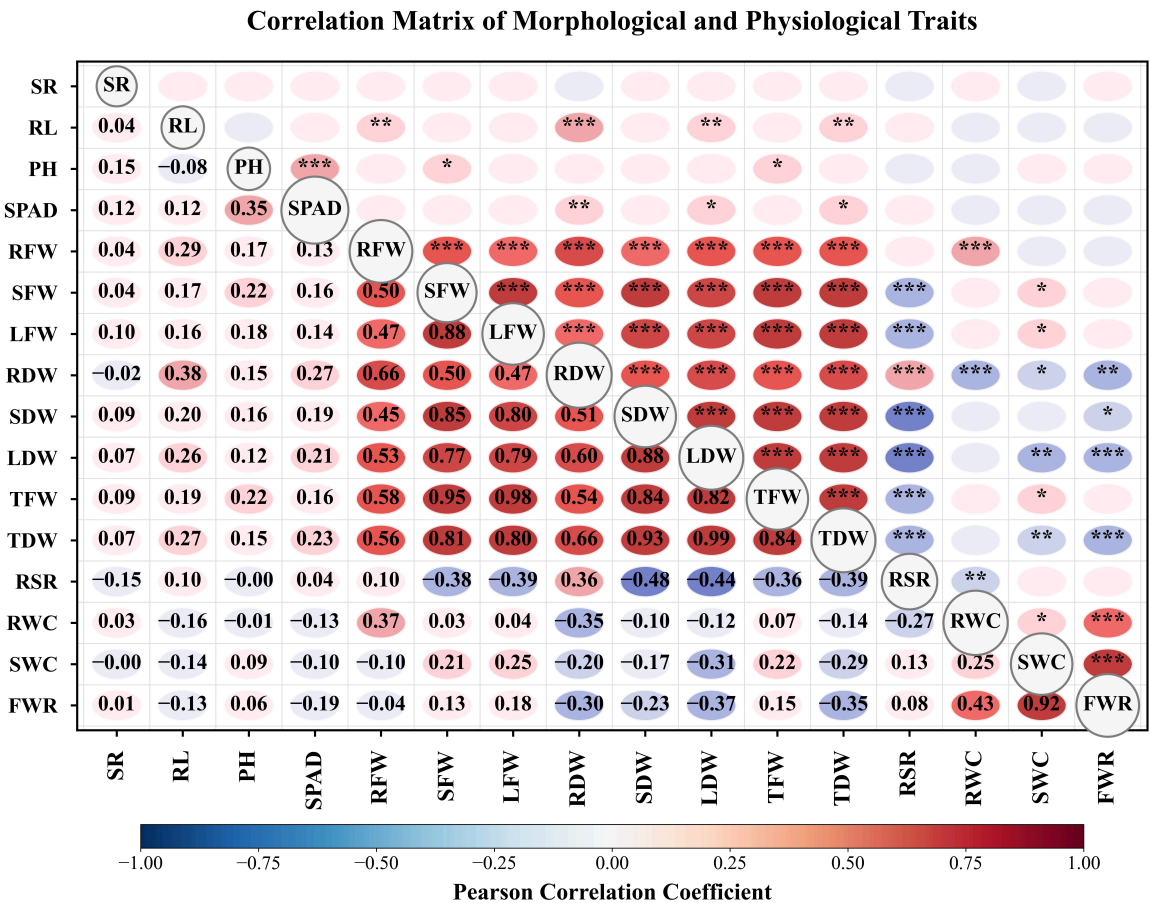


Figure 2. Correlation matrix of 16 salt tolerance-related traits in soybean. Pairwise Pearson correlation coefficients among 16 salt tolerance-related traits. Color intensity indicates the strength of the correlation, with red representing positive correlations and blue representing negative correlations. Asterisks denote significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Diagonal elements are labeled with trait abbreviations.

RSR exhibited moderate negative correlations with SFW ($r = -0.38$), SDW ($r = -0.48$), and LDW ($r = -0.44$), and a moderate positive correlation with RDW ($r = 0.36$) (all $p < 0.001$). This pattern is consistent with a dry matter allocation trade-off between aboveground and belowground tissues, although this interpretation should be confirmed by direct biomass partitioning analyses. SR was weakly correlated with all other indicators ($|r| < 0.15$), providing largely independent information. Collectively, these results indicate that biomass-related traits carry substantial information redundancy and are well-suited for dimensionality reduction via PCA to extract a composite biomass factor, thereby streamlining the salt tolerance evaluation system.

3.3. Construction of a Comprehensive Evaluation Model for Soybean Salt Tolerance Based on Principal Component Analysis

PCA extracted five principal components with eigenvalues > 1 (Figure 3A), cumulatively accounting for 82.90% of the total phenotypic variance (Figure 3B; Supplementary Table S4). PC1 (eigenvalue = 6.47, 40.04% of variance) was driven primarily by biomass-related traits (TFW, SFW, LFW, TDW, SDW, LDW), representing a biomass accumulation dimension that reflected overall growth vigor under salt stress as the dominant evaluation axis. PC2 (eigenvalue = 2.67, 16.55%) was associated with water-related traits (SWC, FWR, RWC), representing a water balance dimension that captured water homeostasis capacity independent of biomass accumulation. PC3 (eigenvalue = 1.75, 10.83%), PC4 (eigenvalue = 1.35,

8.36%), and PC5 (eigenvalue = 1.15, 7.12%) reflected photosynthetic capacity (SPAD), dry matter allocation (RSR), and root morphology (RL), respectively, serving as supplementary dimensions.

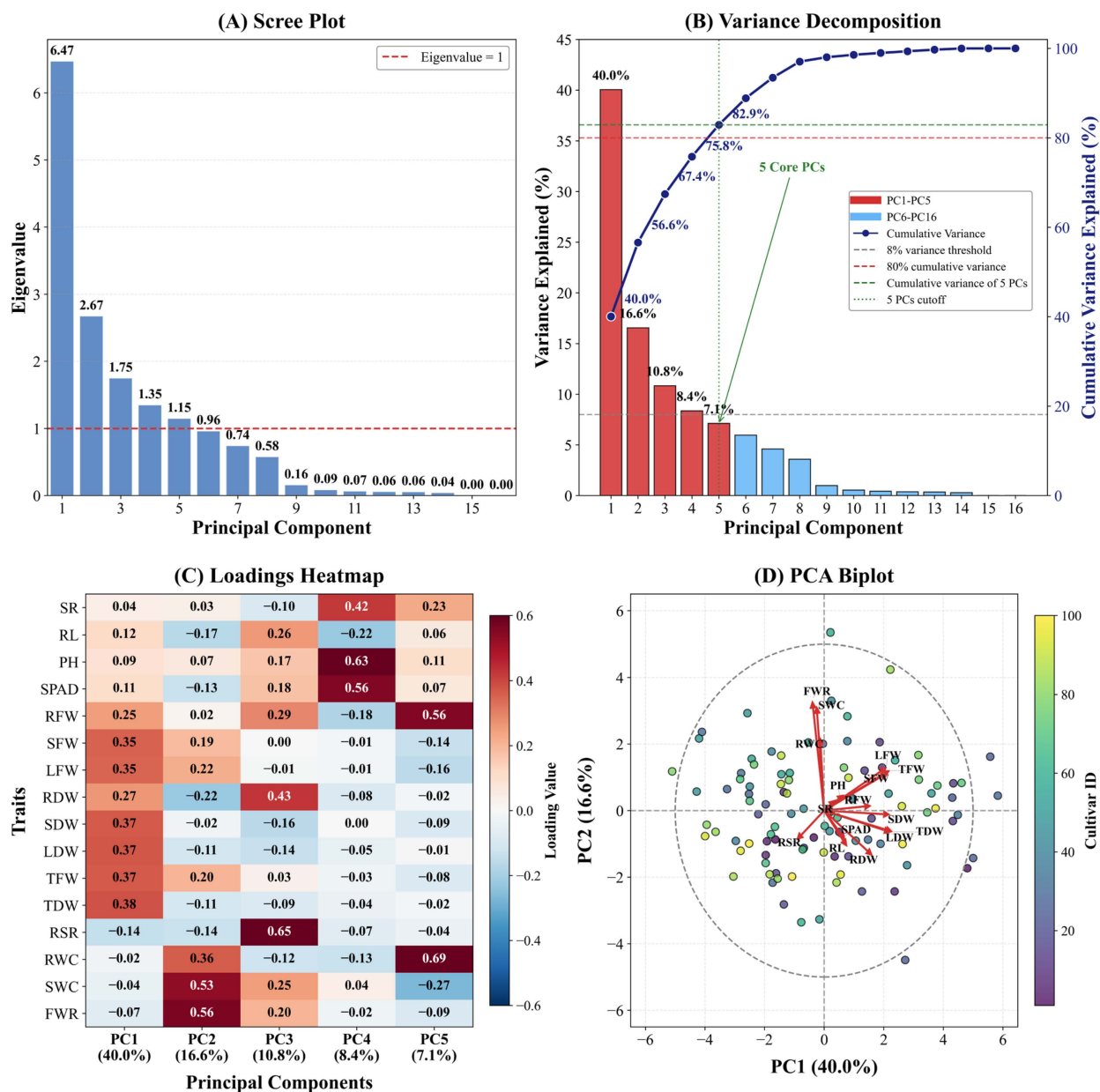


Figure 3. Principal Component Analysis (PCA) of 16 salt tolerance-related traits. (A) Scree plot showing the eigenvalues of each principal component. The dashed line indicates the eigenvalue = 1 criterion. The first five principal components (PC1–PC5) with eigenvalues > 1 were retained. (B) Variance explained by each principal component (bars) and cumulative variance explained (line plot). PC1–PC5 together accounted for 82.90% of the total phenotypic variation. (C) Loading matrix heatmap of the first five principal components. Color intensity represents the magnitude and direction of trait loadings on each component. (D) PCA biplot showing sample scores and trait loadings on PC1 and PC2. Points represent individual accessions, and vectors represent trait loading directions and magnitudes.

Loading heatmaps (Figure 3C) and biplots (Figure 3D) confirmed strong collinearity among biomass indicators and the relative independence of water-related traits. The 100 accessions were broadly distributed across the PC1–PC2 space without distinct clustering, reflecting rich phenotypic diversity within the germplasm panel. Complete PCA results, including

factor loadings (Supplementary Table S5), score coefficients (Supplementary Table S6), and accession-level principal component scores, are provided in the Supplementary Materials. A comprehensive evaluation model was constructed using the five principal components, weighted by their variance contributions, to calculate a composite salt tolerance score for each accession.

3.4. Determination of Optimal Cluster Number and Population Structure Analysis of Soybean Germplasm

The elbow method and silhouette coefficient analysis jointly identified $K = 4$ as the optimal cluster number (Figure 4). The silhouette score at $K = 4$ was 0.22. While below the ideal threshold, this is expected for salt tolerance as a continuously distributed quantitative trait for which discrete class boundaries do not inherently exist. Supporting this choice, the Calinski–Harabasz index (34.12) and Davies–Bouldin index (1.45) each favored $K = 4$ (Supplementary Table S7), and bootstrap resampling (100 iterations) yielded a mean pairwise co-occurrence probability of 0.54 (Supplementary Table S8), confirming moderate stability of cluster assignments across resampling. Collectively, these validity metrics support $K = 4$ as a reasonable partition.

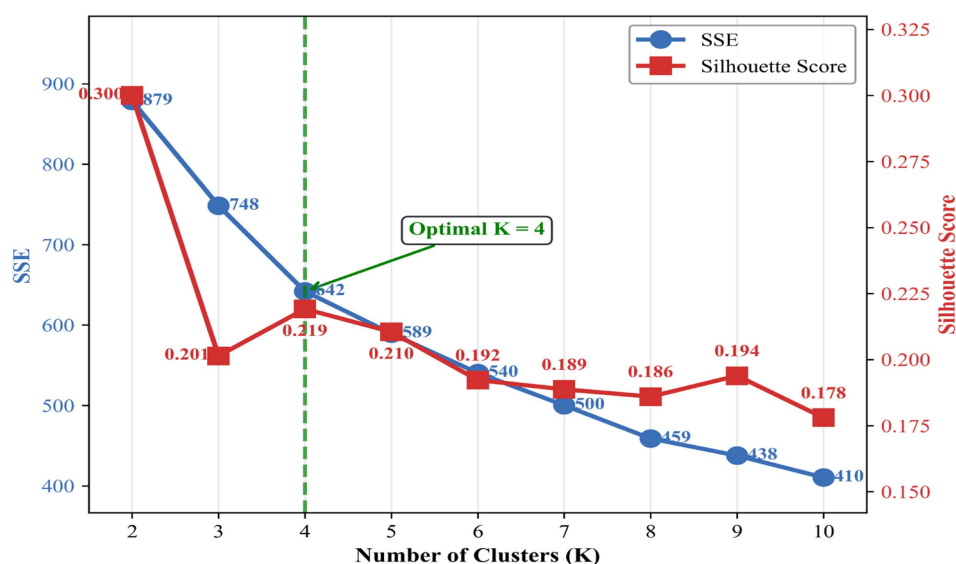


Figure 4. Selection of the optimal number of clusters (K). The optimal number of clusters was determined by combining the elbow method and silhouette coefficient analysis. The blue line represents the sum of squared errors (SSE), which shows a clear inflection point at $K = 4$. The red line represents the silhouette coefficient, which reaches a local maximum at $K = 4$. Based on both criteria, $K = 4$ was selected as the optimal cluster number.

K-means clustering was performed on the five standardized principal component scores derived from the 16-trait PCA. In this procedure, each principal component was implicitly weighted by its variance contribution, as components with larger eigenvalues exerted proportionally greater influence on Euclidean distance calculations. This approach preserves the relative importance of each component as determined by the original data structure.

Clustering classified the 100 accessions into four salt tolerance categories: Class I (highly salt-tolerant, $n = 20$), Class II (moderately salt-tolerant, $n = 16$), Class III (moderately salt-sensitive, $n = 32$), and Class IV (salt-sensitive, $n = 32$). PCA scatter plots revealed clear spatial separation of the four categories along PC1 and PC2 (Figure 5A). Class I accessions were predominantly distributed in the positive region of both axes, reflecting strong biomass accumulation and stable water balance under salt stress. Class IV accessions

clustered in the negative region of both PC1 and PC2, indicating relatively lower biomass accumulation and water content compared with other classes under salt stress. These interpretations are supported by the post hoc trait comparisons in Supplementary Table S9, which confirmed statistically significant differences among classes for biomass-related and water-related traits. Boxplots of composite scores confirmed statistically significant differences among the four categories ($p < 0.001$), with a consistent salt tolerance gradient: Class I > Class II > Class III > Class IV (Figure 5B).

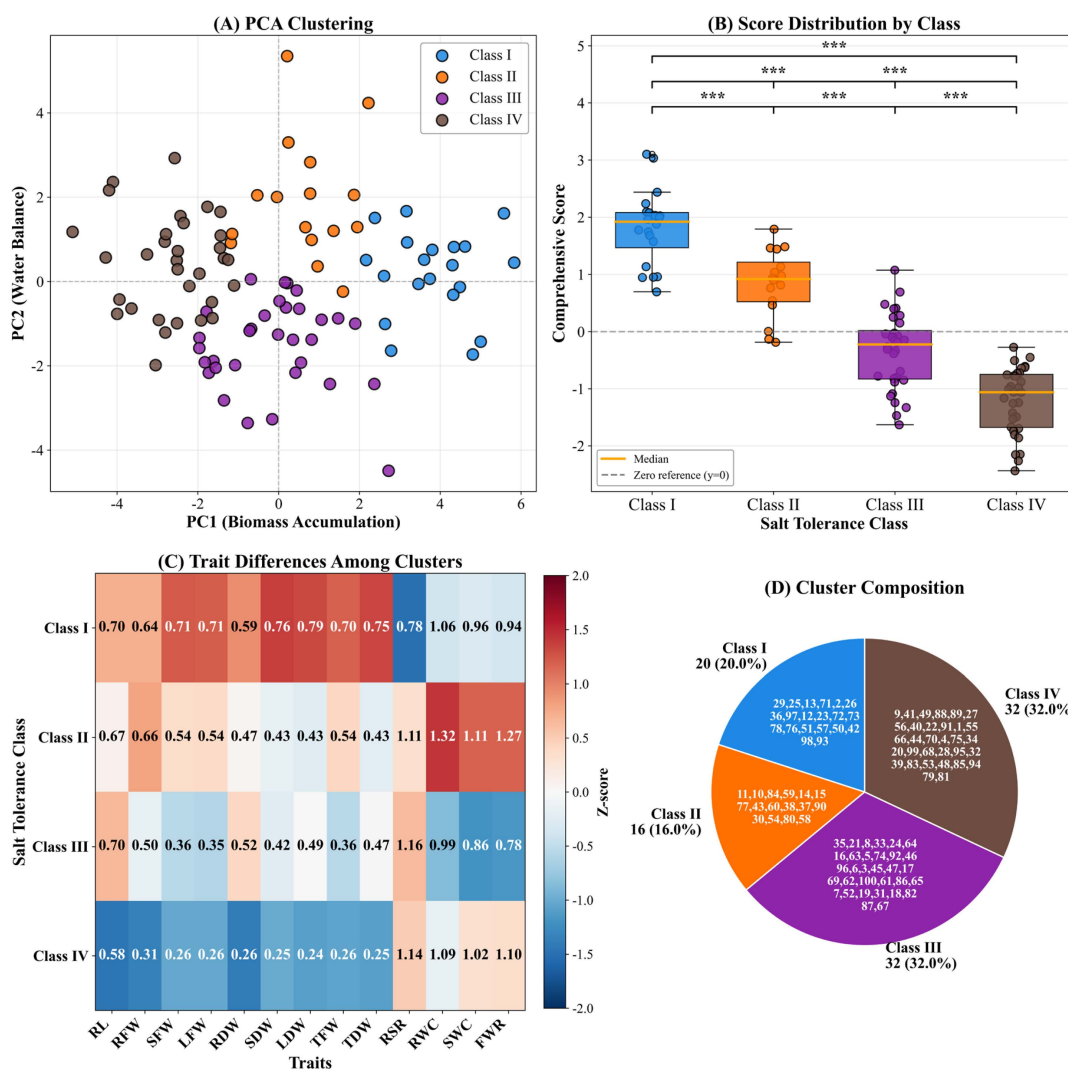


Figure 5. Cluster analysis of soybean accessions based on salt tolerance traits. (A) PCA scatter plot showing the spatial distribution of four salt tolerance classes along PC1 (Biomass Accumulation Factor) and PC2 (Water Balance Factor). (B) Boxplot of comprehensive salt tolerance scores across the four classes. *** $p < 0.001$. (C) Heatmap of standardized mean values of key agronomic traits among the four clusters. Numbers in cells represent raw mean values, and colors indicate Z-scores (red: above mean; blue: below mean). (D) Pie chart showing the composition of the four classes, including accession numbers, proportions, and cultivar IDs. Class I to Class IV represent high, medium–high, medium–low, and low salt tolerance, respectively.

Heatmap analysis further characterized trait differences across categories (Figure 5C). Class I exhibited markedly higher biomass trait values and a lower root-to-shoot ratio, consistent with an efficient biomass accumulation strategy. Class II maintained relatively high water content and fresh-to-dry weight ratio. In contrast, Class IV displayed the lowest biomass trait values, and its water-related traits did not translate into improved growth performance, resulting in pronounced growth suppression under salt stress. One-way

ANOVA with Tukey HSD post hoc comparisons confirmed that biomass-related traits (TFW, SFW, LFW, TDW, SDW, LDW) and water-related traits (SWC, FWR) showed the strongest discrimination among categories (5–6 significant pairwise comparisons), whereas morphological (PH, SR) and physiological (SPAD) traits showed no statistically significant differences among categories (Supplementary Table S9). These post hoc comparisons provide statistical support for the biological interpretation of the clusters that is independent of the clustering procedure itself.

Class distribution analysis revealed that 20.0% of accessions were classified as Class I, while 64.0% fell into the moderately salt-sensitive or salt-sensitive categories (Class III and IV), indicating that highly salt-tolerant accessions are relatively scarce within this germplasm panel (Figure 5D). These results quantitatively capture the range of salt tolerance variation, with Class I accessions representing promising parental resources for further breeding evaluation.

3.5. Identification of Core Salt Tolerance Indicators in Soybean Based on Random Forest and SHAP Analysis

To complement the inherent limitation of linear PCA weighting—which assumes each trait contributes independently and proportionally to the composite score—Random Forest and SHAP analyses were employed to quantify the nonlinear and non-proportional contributions of individual traits. In this analytical design, the PCA-derived comprehensive score served as the response variable, not as an independent external outcome to be predicted, but as a standardized internal metric representing the integrated salt tolerance level of each accession. The objective of the machine learning analysis was therefore to calibrate trait weights within the established PCA evaluation framework, identifying which traits exerted the greatest influence on the composite score and revealing their directional associations. This approach addresses the question of “which traits matter most” in the PCA-based scoring system, rather than building a standalone predictive model for an externally defined phenotype.

Random Forest and SHAP analysis jointly identified key predictors of the comprehensive salt tolerance score (Figure 6). In the Random Forest model, fresh biomass-related traits accounted for the majority of total feature importance: TFW (59.5%), SFW (28.6%), and LFW (4.8%) together contributed 92.9% of the total importance scores, indicating their predominant role in predicting the comprehensive score. The remaining 13 traits each had importance scores below 1.5%. SHAP analysis independently confirmed this ranking: TFW (mean $|SHAP|$ = 0.6440), SFW (0.3241), and LFW (0.0992) were the top three influential features, with all other traits showing SHAP values below 0.05. The concordance between these two independent methods supports the selection of TFW, SFW, and LFW as core indicators for the simplified salt tolerance evaluation model. It should be noted that these feature importance values represent the relative contribution of each trait to model predictive performance, not the proportion of phenotypic variance explained.

SHAP bee swarm and dependence plots further clarified the directional associations of these traits with the salt tolerance score (Figure 7A,B). TFW and SFW exhibited stable positive associations: higher trait values corresponded to positive SHAP values, whereas lower values yielded negative SHAP values. LFW showed a similar but weaker positive trend. Under the experimental conditions examined, these results indicate that fresh biomass accumulation is the phenotypic characteristic most strongly associated with the comprehensive salt tolerance score, with TFW and SFW serving as the preferred indicators for simplified evaluation.

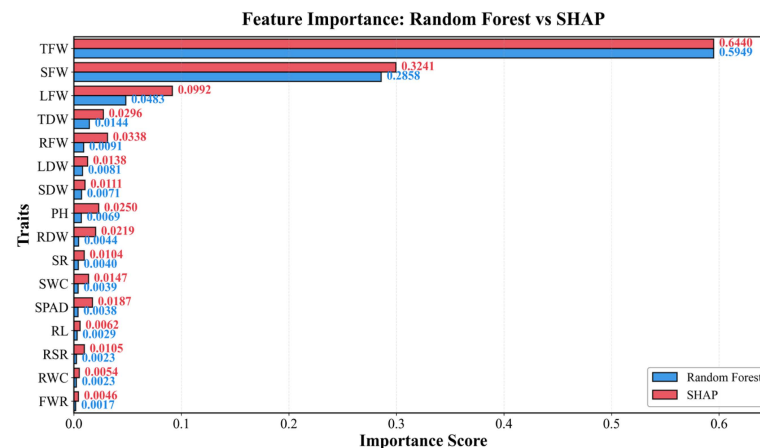


Figure 6. Comparison of feature importance derived from Random Forest and SHAP analysis. Blue bars represent feature importance scores obtained from the Random Forest model, while red bars indicate the mean absolute SHAP (SHapley Additive exPlanations) values. Traits are ordered in ascending order of Random Forest importance. Total fresh weight (TFW), stem fresh weight (SFW), and leaf fresh weight (LFW) consistently show the highest importance across both analyses, identifying them as the most critical traits for predicting the comprehensive salt tolerance score. The strong agreement between the two independent approaches supports the robustness of the key trait selection.

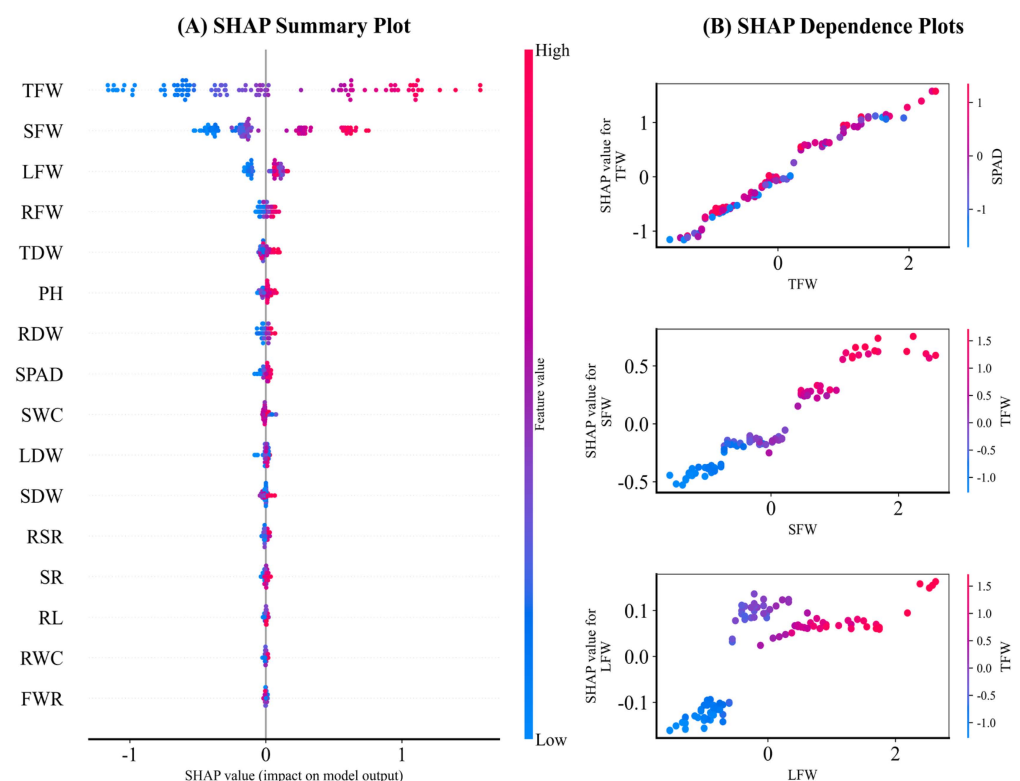


Figure 7. SHAP interpretation of the soybean salt tolerance model. (A) SHAP summary plot showing the contribution of each trait to the comprehensive salt tolerance score. Each dot represents an individual sample, colored by the feature value (blue: low; red: high). TFW, SFW, and LFW exhibit the widest SHAP value distributions, indicating the highest contributions to model predictions. Higher trait values tend to correspond to positive SHAP values, suggesting a positive association with the comprehensive salt tolerance score. (B) SHAP dependence plots for the three key traits. The SHAP values of TFW, SFW, and LFW increase with their respective trait values, showing a clear positive trend. These results confirm the positive associations of these three traits with the comprehensive salt tolerance score.

3.6. Construction and Performance of a Simplified Salt-Tolerance Evaluation Model Based on Core Biomass Traits

3.6.1. Model Construction

Random Forest and SHAP analyses (Section 3.5) identified total fresh weight (TFW), stem fresh weight (SFW), and leaf fresh weight (LFW) as the core phenotypic predictors of the comprehensive salt tolerance score. A parsimonious evaluation model was constructed by applying principal component analysis (PCA) to these three traits. Prior to dimensionality reduction, raw phenotypic values were standardized through Z-score transformation to eliminate dimensional discrepancies. The first principal component accounted for 95.66% of the total variance, indicating that a single dimension sufficiently captured the integrated information carried by the three core traits. The resulting simplified model is expressed as

$$\text{Simplified Score} = 0.359249 \times \text{TFW} + 0.322816 \times \text{SFW} + 0.317936 \times \text{LFW} - 3.765645$$

The coefficients sum to unity, with TFW contributing the largest weight (35.92%), followed by SFW (32.28%) and LFW (31.79%). Detailed model parameters are provided in Supplementary Table S10.

3.6.2. Model Validation and Threshold Optimization

Twenty accessions classified as Class I (highly salt-tolerant) via K-means clustering (Section 3.4) served as the positive reference set for model validation. The simplified model scores correlated strongly with the full 16-trait model scores (Pearson's $r = 0.9745$, Spearman's $\rho = 0.9789$, $p < 0.001$; Figure 8A), confirming that the three-trait model faithfully recapitulates the rankings produced by the full model.

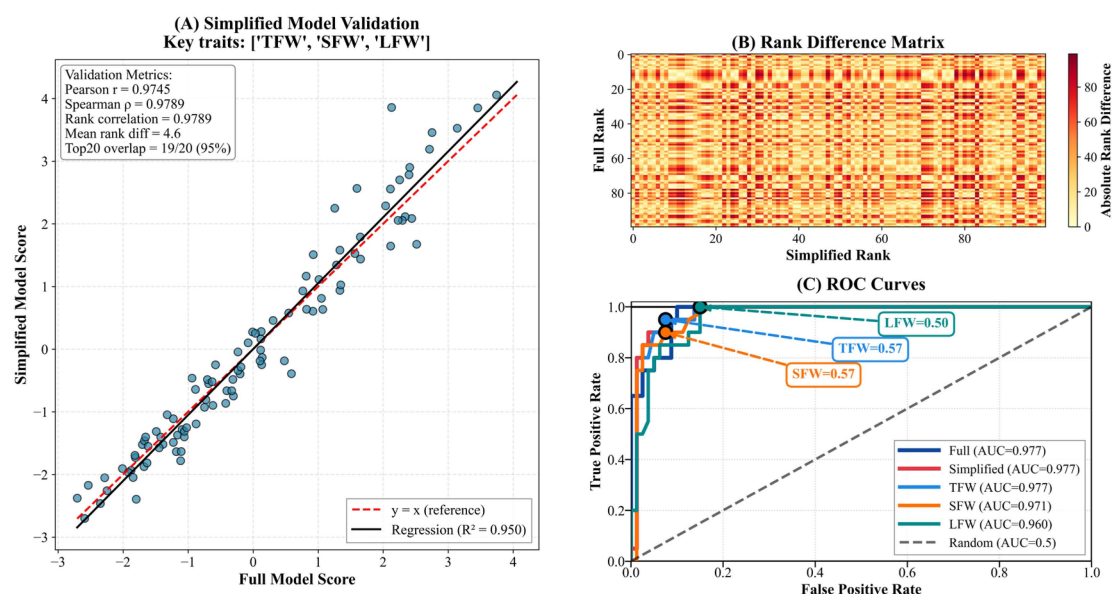


Figure 8. Validation of the simplified salt tolerance evaluation model. (A) Scatter plot of the full model score versus the simplified model score. The red dashed line represents the 1:1 reference line, and the solid black line represents the linear regression fit. Validation metrics (Pearson r , Spearman ρ , rank correlation, mean rank difference, and top-20 overlap) are provided in the inset box. (B) Rank difference matrix showing the absolute difference between the rankings from the full model and the simplified model. Color intensity indicates the magnitude of the rank difference. (C) Receiver operating characteristic (ROC) curves comparing the classification performance of the full model, simplified model, and individual key traits. The area under the curve (AUC) is reported for each curve, and the optimal threshold for each key trait is annotated. The dashed diagonal line represents the performance of a random classifier (AUC = 0.5).

Rank difference matrix analysis further verified ranking stability across the population (Figure 8B). The mean absolute rank shift was 4.6 positions, with no rank reversals observed in the extreme tails of the distribution. The simplified and full models shared 19 of their top 20 accessions (95% overlap). Concordance with K-means-defined Class I reached 85% (17/20) for the simplified model and 80% (16/20) for the full model (Supplementary Table S11).

These minor discrepancies reflect inherent methodological differences among the three approaches. The full model integrates 16 phenotypic traits, whereas the simplified model focuses exclusively on three core biomass indicators. K-means clustering groups accessions by overall phenotypic similarity, while PCA-based scoring ranks them along a continuous tolerance axis. All pairwise concordance rates exceeded 80%, providing convergent validation of the core trait selection.

ROC analysis, with Class I (defined from the internal composite score distribution) as the positive class, yielded an AUC of 0.970 for the simplified model, approaching that of the full model (AUC = 1.000; Figure 8C). These ROC/AUC results should be interpreted as internal consistency metrics that evaluate the agreement between the simplified model and the full-index framework, rather than as validation against an independent biological standard, because the positive class was defined using an internal scoring criterion. Optimal single-trait thresholds determined by maximizing the Youden index were: TFW > 0.57 (AUC = 0.977), SFW > 0.57 (AUC = 0.971), and LFW > 0.50 (AUC = 0.960) (Supplementary Table S12). A two-of-three joint decision rule was established to facilitate practical screening: an accession is classified as salt-tolerant when at least two of the three traits exceed their respective thresholds. This rule achieved 94.0% accuracy, 95.0% sensitivity, and 93.8% specificity (Supplementary Table S13).

3.6.3. Summary of Model Performance

The three-trait simplified model reproduces the rankings of the full 16-trait model (Pearson's $r = 0.9745$, top-20 overlap = 95%) and aligns closely with unsupervised phenotypic classification (Class I concordance = 85%). With defined single-trait thresholds and a validated joint decision rule (94.0% accuracy), this parsimonious model enables efficient preliminary screening of soybean germplasm for seedling-stage salt tolerance, substantially reducing phenotyping effort. Its broader applicability requires further validation across diverse environments and genetic backgrounds.

3.7. Stability, Generalizability, and Biological Coherence of Core Evaluation Indicators

3.7.1. Nested Cross-Validation Design

A 10-fold nested cross-validation framework was implemented to evaluate the stability and generalizability of the simplified model. The nested design comprised two independent loops: the outer 10-fold loop generated unbiased performance estimates on held-out test subsets, while the inner 10-fold loop performed hyperparameter optimization exclusively on training data. A fixed global random seed (SEED = 42) was applied throughout to ensure full reproducibility.

A strict training-test isolation strategy was enforced for model optimization and validation processes to reduce data leakage risk and improve model stability. However, since the three core traits (TFW, SFW, and LFW) were fixed as model inputs based on full-dataset feature selection without secondary screening in cross-validation folds, the validation results cannot entirely exclude data leakage and overfitting risks, and primarily represent the internal robustness of the simplified model. All preprocessing and parameter calibration procedures—including Z-score standardization, PCA transformation, and Youden index-based threshold optimization—were independently fitted within each outer training fold and never incorporated test-fold information. The three core traits (TFW,

SFW, LFW) were fixed as model inputs based on full-dataset feature selection and were not re-screened across folds. Inner-loop optimization screened PCA component number (1, 2, or 3) and standardization scheme (Z-score enabled or disabled), with Pearson's correlation as the selection criterion. The combination of two PCA components with Z-score standardization was stably selected across all training subsets (inner-fold Pearson correlations: 0.9673–0.9856; Supplementary Table S14).

3.7.2. Cross-Fold Stability of Model and Core Traits

The simplified model exhibited stable performance across all outer test folds (Supplementary Table S15). Regression metrics confirmed consistent agreement with the full model: Pearson $r = 0.979 \pm 0.016$, Spearman $\rho = 0.965 \pm 0.037$, $R^2 = 0.935 \pm 0.048$, RMSE = 0.358 ± 0.143 , and MAE = 0.288 ± 0.103 (Figure 9A). Classification metrics further supported robust discrimination: AUC = 0.975 ± 0.050 , accuracy = 0.900 ± 0.063 , sensitivity = 0.917 ± 0.171 , and specificity = 0.901 ± 0.075 . Bootstrap resampling (1000 iterations) yielded 95% confidence intervals of [0.944, 0.998] for AUC, [0.840, 0.960] for accuracy, [0.750, 1.000] for sensitivity, and [0.829, 0.963] for specificity. The wider confidence interval for sensitivity reflects the limited number of Class I accessions ($n = 20$), an inherent constraint of the natural population that may be addressed by expanding sample size in future work. Per-fold performance metrics are provided in Supplementary Table S16.

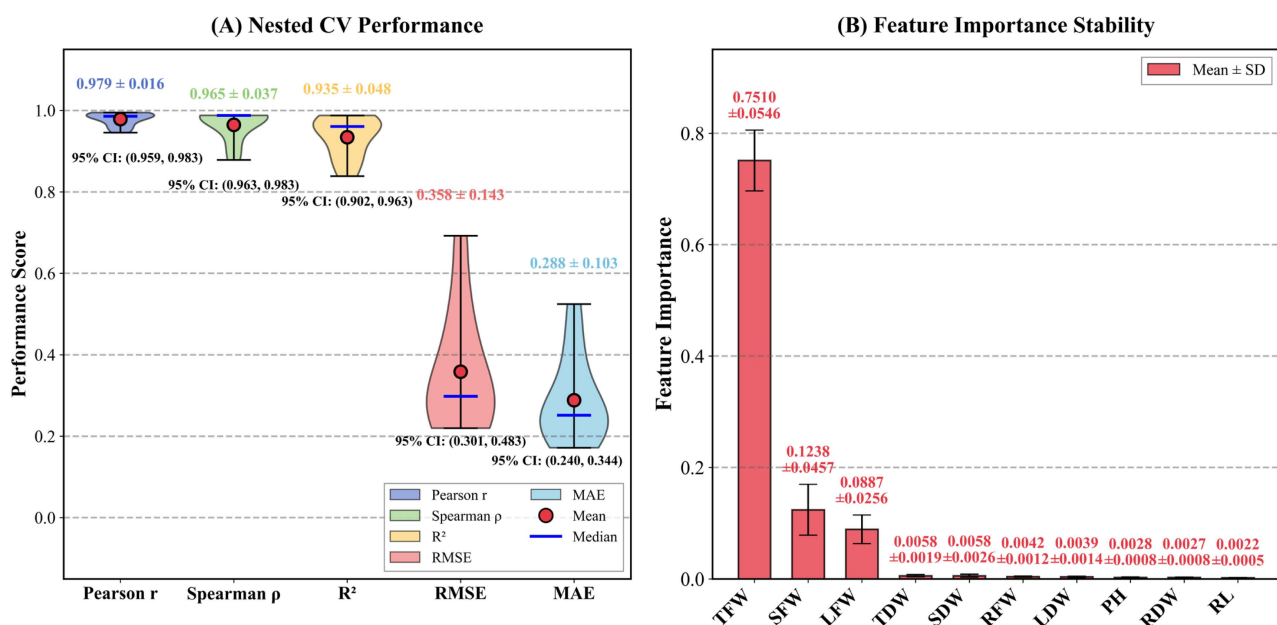


Figure 9. 10-fold nested cross-validation and feature importance stability analysis. (A) Nested CV performance. Violin plots show the distribution of five regression metrics across 10 folds. Red dots indicate mean values. Values are presented as mean $\pm$ SD with 95% confidence intervals. (B) Stability analysis of feature importance. Bar plots show the mean feature importance (red bars) $\pm$ standard deviation (error bars) for each predictor variable across cross-validation runs.

At the individual trait level, TFW, SFW, and LFW maintained consistently high cross-validation scores across all folds (means of 0.9785, 0.9661, and 0.9256, respectively) with minimal inter-fold variation. Feature importance rankings remained invariant under repeated random partitioning: TFW consistently dominated (importance = 0.7510), followed by SFW (0.1238) and LFW (0.0887). This stable hierarchy across independent data subsets excludes sampling artifacts and overfitting as explanations for the observed importance pattern (Figure 9B).

3.7.3. Biological Coherence of Core Evaluation Indicators

Correlation and network analyses revealed close coordination among the three core fresh-weight traits (Figure 10). Pairwise Pearson correlations among TFW, SFW, and LFW ranged from 0.88 to 0.98, indicating highly synergistic variation in seedling fresh biomass under salt stress. These fresh-weight traits also exhibited strong positive correlations with dry-weight biomass indices (TDW, SDW, LDW; $r = 0.66\text{--}0.99$), demonstrating that fresh biomass variation largely captures whole-plant growth responses to salt stress. In contrast, morphological traits such as plant height showed weak correlations with the core biomass indicators, further supporting biomass accumulation as a more sensitive phenotypic indicator of seedling salt tolerance under the examined conditions.

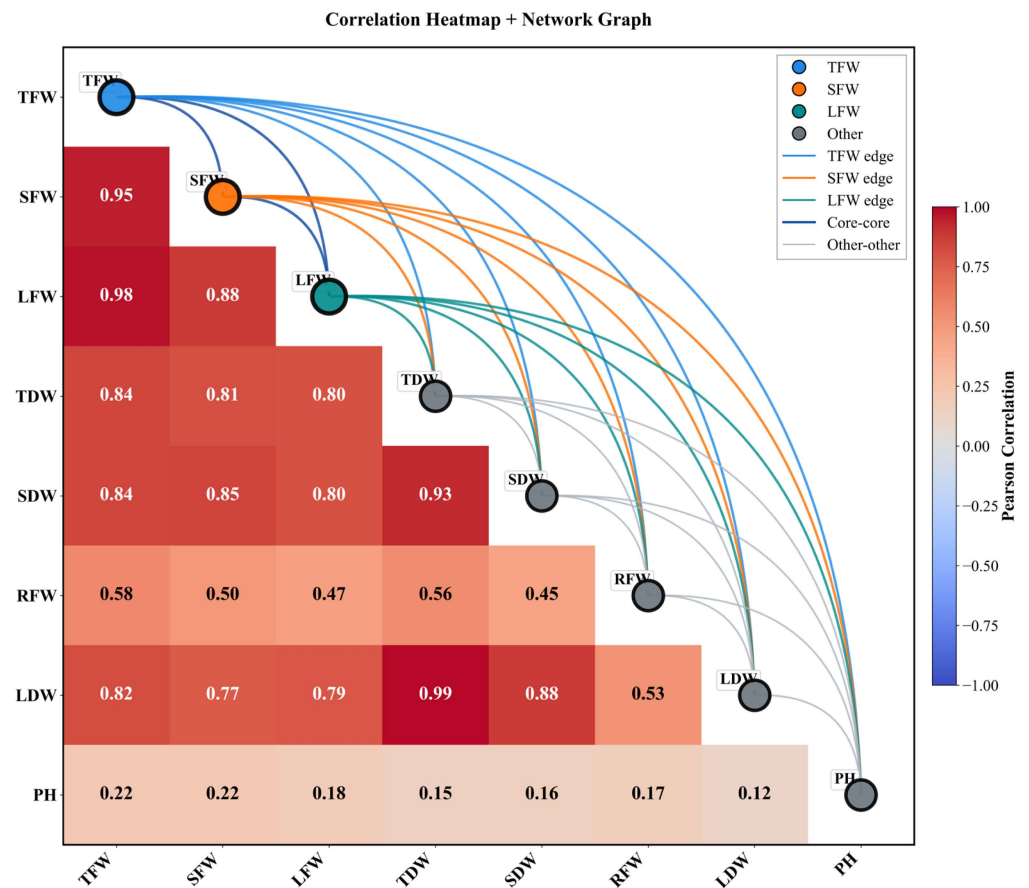


Figure 10. Correlation heatmap and network graph of key phenotypic traits for soybean salt tolerance. The lower-left triangle presents a Pearson correlation heatmap, and the upper-right triangle displays an association network graph. In the heatmap, color intensity and numerical values represent correlation coefficients, with red indicating stronger positive correlations. In the network graph, node colors distinguish core traits (TFW, SFW, LFW) from other traits, while edges illustrate the strength of associations between them. TFW, SFW, and LFW are strongly positively correlated with each other ($r = 0.88\text{--}0.98$) and exhibit high coordination with other fresh and dry weight-related traits, indicating that biomass-related traits constitute a core group of indicators for salt tolerance response in soybean.

Collectively, the convergent evidence from nested cross-validation stability, bootstrap-derived confidence bounds, and biologically interpretable trait coordination indicates that TFW, SFW, and LFW constitute a robust and parsimonious indicator set for simplified salt tolerance evaluation. The consistent cross-fold performance suggests that the simplified model avoids substantial overfitting and maintains reliable performance across independent data partitions within the studied germplasm panel.

4. Discussion

4.1. Fresh Biomass Accumulation as the Primary Determinant of Soybean Salt Tolerance Variation

Fresh biomass accumulation emerged as the phenotypic characteristic most strongly associated with salt tolerance variation in soybean seedlings across the examined germplasm panel within the Random Forest model framework. Machine learning-based phenotypic screening demonstrated that three biomass-related traits—total fresh weight (TFW), stem fresh weight (SFW), and leaf fresh weight (LFW)—collectively accounted for 92.9% of total feature importance in the Random Forest model predictions, while the remaining 13 traits each contributed less than 1.5% to the model's predictive performance. It is worth noting that these importance values reflect the relative contribution of each trait to the model under the specific analytical conditions of this study, and are not direct evidence of biological causation or a universal measure of physiological importance. Nevertheless, the observed importance pattern is physiologically interpretable: salt stress triggers simultaneous osmotic perturbation and ionic toxicity, which suppress cellular elongation and proliferation, ultimately constraining seedling growth and reducing fresh biomass accumulation under saline conditions [7,17].

Previous studies have shown that salt-tolerant soybean genotypes can maintain higher fresh biomass through coordinated regulation of osmotic adjustment, ion homeostasis, and root–microbe interactions [23–28]. The transcription factors GmDREB7 and GmCOL1a promote proline biosynthesis by activating GmP5CS expression, maintaining cellular turgor and alleviating salt-induced growth repression [24,25]. Enhanced expression of BADH improves cellular osmotic balance and preserves biomass productivity under salt stress [26]. Plasma membrane and tonoplast Na^+/H^+ antiporters (SOS1 and NHX1) further mitigate ionic damage by mediating Na^+ efflux and vacuolar sequestration, stabilizing cellular K^+/Na^+ homeostasis [27,28]. Whether these specific mechanisms operate in the accessions examined here remains to be experimentally verified.

Fresh biomass traits exhibited greater sensitivity and discriminatory capacity for salt tolerance than dry biomass indices, consistent with previous crop stress studies [29]. This functional divergence likely arises from distinct physiological response kinetics: fresh weight dynamically reflects short-term cellular water status and turgor stability under stress, whereas dry weight integrates long-term carbon assimilation and exhibits delayed responses. The utility of fresh biomass as a salt tolerance indicator has been reported across major cereal crops, including maize, wheat, and rice [30–32]. Among all traits evaluated here, TFW displayed strong discriminatory performance ($\text{AUC} = 0.977$), supporting its value as a core indicator for quantifying soybean seedling salt adaptation under the examined conditions.

The strong phenotypic correlations among TFW, SFW, and LFW can redistribute feature importance across these variables, potentially biasing single-model screening approaches. To address this, this study integrated two complementary analytical frameworks—Random Forest ranking and SHAP interpretative analysis—both of which consistently prioritized the same three fresh biomass traits. Cross-validation further confirmed the stability of this trait selection, with consistent feature importance rankings across independent data partitions. This multi-layered validation strategy supports that the identified core indicators reflect meaningful stress-adaptive phenotypic variation rather than statistical artifacts arising from trait collinearity.

Soybean accessions inherently differ in seedling size and developmental rate, which may cause baseline disparities in absolute biomass values and potentially confound salt tolerance evaluation. To address this, two strategies were implemented. First, the salt tolerance coefficient (STC), calculated as the trait value ratio between salt-stressed and control conditions for each genotype, was adopted as the core evaluation index. This

ratio-based normalization removes the influence of intrinsic basal growth differences; consequently, a genotype with inherently larger seedlings but weak salt adaptability will still exhibit a low STC and be correctly classified as salt-sensitive. Second, two-way ANOVA revealed statistically significant accession $\times$ treatment interactions for all 16 traits ($p < 0.01$), providing statistical evidence that salt stress imposes heterogeneous inhibitory effects on different accessions. These results indicate that the observed STC variation reflects genuine differences in salt stress sensitivity among accessions, rather than proportional amplification of inherent growth disparities.

4.2. Limited Predictive Utility of SPAD Value and Root-to-Shoot Ratio

SPAD value and root-to-shoot ratio (RSR) each accounted for less than 1.5% of total feature importance, consistent with prior soybean salt stress research showing that chlorophyll-related traits are not primary predictors of salt adaptation in soybean [6]. In contrast, SPAD serves as an effective screening indicator in rice and wheat [33,34], suggesting species-specific stress response patterns. Such interspecific discrepancies may reflect inherent differences in chlorophyll stability under ionic toxicity, the moderate stress intensity employed here, or the reliance of earlier studies on single-indicator analyses that do not account for inter-trait redundancy.

Although RSR is a well-documented adaptive trait in drought tolerance research, it contributed minimally to predicting the salt tolerance score in this study. Root architectural plasticity and biomass reallocation to roots primarily enhance water acquisition under water deficit conditions [35,36]. Under salt stress, ionic toxicity predominantly impairs aboveground seedling growth, limiting adaptive variation in root morphology and RSR. These findings suggest that SPAD and RSR exhibit stress-type-specific functional relevance, with limited universal applicability for assessing soybean salt tolerance under the conditions examined.

4.3. Functional Interpretation of Core Biomass Traits: Phenotypic Patterns and Breeding Implications

SHAP-based analysis revealed consistent positive, monotonic associations between core fresh biomass traits and comprehensive salt tolerance scores. Elevated TFW and SFW values corresponded to higher SHAP values without observable saturation, suggesting that biomass accumulation is positively associated with salt tolerance across the range examined. Clear separation between salt-tolerant and salt-sensitive accessions in SHAP distributions further supports the utility of fresh biomass traits for distinguishing differential salt tolerance responses. It should be emphasized that SHAP values reflect statistical associations within the predictive model and do not establish causal physiological relationships.

Fresh biomass-based salt tolerance is a genetically complex quantitative trait. Previously identified QTLs associated with shoot fresh weight under salt stress on soybean chromosomes 6, 15, and 19 provide genetic evidence for the heritability of biomass-related salt adaptation [37]. The three core traits exhibited hierarchical discriminatory capacity: TFW (AUC = 0.977) outperformed SFW (AUC = 0.971) and LFW (AUC = 0.960), likely reflecting its ability to capture integrated whole-plant growth responses. The utility of fresh biomass traits for crop salt tolerance evaluation has been reported across multiple species [38–40]. The empirically optimized phenotypic thresholds established in this study (TFW > 0.57, SFW > 0.57, LFW > 0.50) provide quantitative screening criteria that may facilitate standardized salt-tolerant germplasm selection, pending validation across diverse environments.

4.4. Methodological Contributions and Breeding Utility of the Simplified Three-Trait Model

It should be clarified that the Random Forest model in this study was not designed as an independent predictive model with a separate biological response variable. Rather, it served as a post hoc weight calibration tool within the established PCA evaluation pipeline. In crop salt tolerance research, PCA composite scores are widely adopted as standardized metrics for germplasm classification, yet their linear construction inherently limits the representation of trait interactions. The machine learning component of this study was therefore introduced to quantify the nonlinear contributions of individual traits to the PCA-derived score, thereby complementing the proportional weighting assumption of PCA. This analytical framework aligns with the standard research paradigm in crop stress tolerance evaluation, where the goal is indicator optimization rather than external outcome prediction. The nested cross-validation procedure further ensured that the calibrated trait rankings were not artifacts of data dependency.

The parsimonious three-trait model achieved predictive accuracy comparable to the full 16-trait evaluation system (AUC = 0.970; nested cross-validation accuracy = 90.0%), while substantially reducing phenotyping effort. This framework offers several methodological features. First, the integration of Random Forest feature selection with SHAP-based interpretation clarifies both the priority and the directional association of core indicators, complementing traditional linear weighting methods. Second, the data-driven trait simplification pipeline reduces subjective selection bias. Third, the multi-dimensional validation framework—encompassing nested cross-validation, threshold optimization, and correlation network analysis—provides convergent assessment of indicator stability.

TFW, SFW, and LFW can be measured non-destructively in the field without specialized equipment. A two-stage screening pipeline is proposed for consideration in breeding programs: primary screening by TFW threshold (STC > 0.57) to exclude salt-sensitive accessions, followed by secondary screening using the multi-trait joint decision rule. This approach eliminates the need for dry weight determination and aligns with hierarchical screening strategies adopted in other crop stress resistance studies [41–43]. However, the operational value of this pipeline requires confirmation through independent field validation.

The simplified model preserved accession rankings with high fidelity relative to the full model (mean rank difference = 4.6; top-20 overlap = 95%), with no rank reversals in the extreme phenotypic groups. This level of rank conservation suggests that the simplified framework may be adequate for preliminary germplasm evaluation, where the primary objective is identification of extreme phenotypes.

4.5. Phenotypic Variation and Cross-Validation Support Core Indicator Stability

The three core fresh biomass traits exhibited substantial phenotypic variation across the 100 diverse soybean accessions (CV = 42.6–48.4%; range = 0.07–0.97), providing a robust empirical basis for trait screening. Independent cross-validation confirmed consistent predictive performance and stable feature importance rankings, with TFW consistently identified as the dominant contributor (importance = 0.7510). This stability across random data partitions supports that the core indicator hierarchy is not an artifact of specific sample composition.

Strong positive correlations among TFW, SFW, and LFW ($r = 0.88$ – 0.98) reflect coordinated whole-plant growth responses under systemic salt stress, where saline toxicity inhibits growth across vegetative tissues [44]. High correlations between fresh and dry biomass indices ($r = 0.66$ – 0.99) indicate that fresh biomass variation largely captures dry biomass information, providing a physiological rationale for the fresh weight-centered simplified evaluation approach.

Two-way ANOVA confirmed statistically significant accession $\times$ treatment interactions ($p < 0.01$) for all 16 traits, demonstrating that phenotypic variation in STC values reflects genuine differences in salt stress responses among accessions, rather than inherent disparities in basal growth potential. This provides the statistical foundation for accession-specific salt tolerance evaluation in this study.

4.6. Study Limitations and Future Research Perspectives

Several limitations of this study should be acknowledged. First, the germplasm panel was restricted to 100 cultivated soybean accessions; extension to wild soybean and exotic germplasm is needed to assess broader applicability. Second, evaluations were conducted under a single moderate salt stress condition (0.3% soil salt content, EC 5.0 dS/m); trait importance hierarchies may differ under other stress intensities. Third, only seedling-stage traits were measured; the relationship between seedling salt tolerance and field yield remains to be established. Fourth, the analysis was conducted at the phenotypic level without direct molecular validation of implicated regulatory pathways. Fifth, the model was trained and internally validated within a single population; external validation using independent germplasm panels is essential before recommending operational use. Sixth, trait collinearity among biomass indices may influence feature importance estimates, and permutation-based methods should be employed in future work. Seventh, the ROC analysis employed an internal scoring criterion (top 20% of composite scores) to define the positive class; the reported AUC values therefore reflect model consistency rather than accuracy against an external biological standard. Future studies should incorporate independent physiological indicators—such as tissue ion content or osmotic adjustment capacity—as reference standards.

Future research directions include: (1) GWAS-based mapping of QTLs associated with TFW and SFW, combined with expression analysis of candidate genes (P5CS, BADH, SOS1, NHX1) in contrasting accessions [45]; (2) multi-omics integration to elucidate pathways sustaining fresh biomass under salt stress [46]; (3) multi-environment field trials across salt concentrations and growing conditions to evaluate genotype-by-environment interactions [47]; and (4) utilization of identified salt-tolerant accessions as parental materials to assess the heritability and breeding value of biomass-based salt tolerance traits in segregating populations [48].

5. Conclusions

This study identified three fresh biomass traits—TFW, SFW, and LFW—as the primary contributors to predicting the comprehensive salt tolerance score, collectively accounting for over 93% of total feature importance in the Random Forest model. The simplified three-trait model showed strong consistency with the full 16-trait model (Pearson $r > 0.970$; AUC = 0.970) and maintained stable performance under nested cross-validation (accuracy = 90.0%; AUC = 0.975 ± 0.050) with bootstrap 95% confidence intervals (AUC [0.944, 0.998]) supporting reliability. Statistically significant accession $\times$ treatment interactions ($p < 0.01$) confirmed that STC variation reflects genuine differences in salt stress responses among accessions. Empirically optimized single-trait thresholds (TFW > 0.57 , SFW > 0.57 , LFW > 0.50) provide quantitative criteria for preliminary screening. These conclusions are based on seedling-stage phenotypic data from a single site, a single growing season, and a single salt concentration (0.3% soil salt content). Multi-environment, multi-population, and multi-year validation is required to establish the broader applicability of this simplified evaluation framework for operational breeding programs.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy16141394/s1>, Table S1. Monthly meteorological condi-

tions during the soybean growing season (June–October 2025) at the experimental site in Dezhou, Shandong Province; Table S2. Detailed information of 100 soybean germplasm accessions used in this study; Table S3. Two-way ANOVA results for 16 agronomic traits at the seedling stage under salt stress and control conditions; Table S4. Eigenvalues, variance proportions, and cumulative variance of the five principal components; Table S5. Factor loading matrix of 16 traits on five principal components; Table S6. Score coefficient matrix of 16 traits on five principal components; Table S7. Cluster validity indices for $K = 2$ to $K = 10$; Table S8. Bootstrap clustering stability analysis; Table S9. One-way ANOVA and Tukey HSD post hoc comparisons of traits among salt tolerance classes; Table S10. Mathematical expression and parameters of the simplified evaluation model; Table S11. Top20 overlap and unique accessions among three classification methods; Table S12. Optimal thresholds and AUC values for single indicators based on the Youden index; Table S13. Performance metrics of the two-indicator joint decision rule; Table S14. Hyperparameter optimization results from inner 10-fold cross-validation; Table S15. Performance of the simplified model evaluated by 10-fold nested cross-validation with bootstrap 95% confidence intervals; Table S16. Performance metrics of the simplified model across 10 outer cross-validation folds.

Author Contributions: Conceptualization, Y.T. and J.Z.; methodology, Y.T. and F.H.; software, Y.T.; validation, Y.T., P.C. and C.L.; formal analysis, C.W. and G.Z.; investigation, Q.G. and F.G.; resources, J.Z.; data curation, Y.T.; writing—original draft preparation, Y.T.; writing—review and editing, J.Z. and F.H.; visualization, Y.T.; supervision, J.Z.; project administration, J.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shandong Agricultural Improved Variety Project (Breeding of New Soybean Varieties with Salt-Tolerance and High Planting Density-Tolerance), grant number 2024LZGC030; the Special Fund for the Construction of National Modern Agricultural Industrial Technology System, grant number CARS-04-CES15; the Shandong Modern Agricultural Industrial Technology System, grant number SDAIT-28-13; and the Internal Research Project of Dezhou Academy of Agricultural Sciences, grant number YNXM2024007. The APC was funded by the corresponding author.

Data Availability Statement: The raw phenotypic data, statistical analysis data, and analytical code generated and analyzed in this study are not publicly available due to ongoing further research and data confidentiality requirements. However, the minimal dataset supporting the findings of this study is available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to express their sincere gratitude to the staff of Dezhou Academy of Agricultural Sciences for their valuable technical support during the field experiments and phenotypic trait determination. We also appreciate the assistance from the laboratory colleagues for their help in data sorting and preliminary analysis. Additionally, we thank the relevant departments for providing experimental sites and necessary experimental materials to ensure the smooth progress of this study. We would also like to thank the editorial department and relevant personnel for their guidance and assistance during the manuscript submission and review process.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of Variance
AUC	Area Under the Curve
BADH	Betaine Aldehyde Dehydrogenase
CV	Coefficient of Variation
EC	Electrical Conductivity
FWR	Fresh-to-Dry Weight Ratio

GWAS	Genome-Wide Association Study
LDW	Leaf Dry Weight
LFW	Leaf Fresh Weight
PC	Principal Component
PCA	Principal Component Analysis
PH	Plant Height
QTL	Quantitative Trait Locus
RDW	Root Dry Weight
RF	Random Forest
RFW	Root Fresh Weight
RL	Root Length
ROC	Receiver Operating Characteristic
RSR	Root-to-Shoot Ratio
RWC	Root Water Content
SD	Standard Deviation
SDW	Stem Dry Weight
SFW	Stem Fresh Weight
SHAP	SHapley Additive exPlanations
SPAD	Soil and Plant Analyzer Development
SR	Survival Rate
STC	Salt Tolerance Coefficient
SWC	Shoot Water Content
TDW	Total Dry Weight
TFW	Total Fresh Weight

References

1. FAO; ITPS. *Status of the World's Soil Resources (SWSR)—Main Report*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2015.
2. Shokri, N.; Hassani, A.; Sahimi, M. Multi-scale soil salinization dynamics from global to pore scale: A review. *Rev. Geophys.* **2024**, *62*, e2023RG000804. [[CrossRef](#)]
3. Guan, R.X.; Guo, X.Y.; Qu, Y.; Zhang, Z.W.; Bao, L.G.; Ye, R.Y.; Chang, R.Z.; Qiu, L.J. Salt tolerance in soybeans: Focus on screening methods and genetics. *Plants* **2023**, *13*, 97. [[CrossRef](#)] [[PubMed](#)]
4. Linh, N.; Cham, L.; Thang, V. Effects of salinity stress on the growth, physiology, and yield of soybean (*Glycine max* (L.) Merrill). *Vietnam J. Agric. Sci.* **2021**, *4*, 1043–1055. [[CrossRef](#)]
5. Pruthi, R.; Chaudhary, C.; Chapagain, S.; Abozaid, M.M.E.; Rana, P.; Kondi, R.K.R.; Fritsche-Neto, R.; Subudhi, P.K. Deciphering the genetic basis of salinity tolerance in a diverse panel of cultivated and wild soybean accessions by genome-wide association mapping. *Theor. Appl. Genet.* **2024**, *137*, 238. [[CrossRef](#)] [[PubMed](#)]
6. Arshad, K.; Wang, Y.; Han, S.; Zheng, M.; Xie, M.; Song, Y.; Hu, L.; Ou, R.; Gu, M.; Ouyang, C.; et al. GWAS on HTP-enabled dynamic traits unravels novel genetic architecture of salt tolerance in soybean. *Plant J.* **2025**, *122*, e70177. [[CrossRef](#)] [[PubMed](#)]
7. Meng, R.; Li, Z.J.; Feng, W.; Chen, Y.; Liu, L.P.; Yang, C.Y.; Lu, X.L.; Wang, X.P. Comprehensive evaluation of salt tolerance at different growth stages of soybean and screening of salt-tolerant germplasm. *Acta Agron. Sin.* **2025**, *51*, 1991–2008. [[CrossRef](#)]
8. Sun, X.J.; Hu, Z.; Jiang, X.M.; Wang, S.J.; Chen, X.Q.; Zhang, H.Y.; Zhang, H.; Jiang, Q.Y. Identification, evaluation and screening of salt-tolerant soybean germplasm resources at seedling stage. *Acta Agron. Sin.* **2024**, *50*, 2179–2186. [[CrossRef](#)]
9. Kokebie, D.; Enyew, A.; Kassaw, E.; Wendu, A.; Worku, G.; Fentie, T. Evaluating salinity tolerance in soybean (*Glycine max* L.) varieties under salinity stress: A multivariate analysis approach using PCA and GGE Biplot models. *J. Plant Interact.* **2025**, *20*, 1. [[CrossRef](#)]
10. Singhal, R.; Izquierdo, P.; Ranaweera, T.; Segura Abá, K.; Brown, B.N.I.; Lehti-Shiu, M.D.; Shiu, S.H. Using supervised machine-learning approaches to understand abiotic stress tolerance and design resilient crops. *Philos. Trans. R. Soc. B Biol. Sci.* **2025**, *380*, 20240252. [[CrossRef](#)] [[PubMed](#)]
11. Rico-Chávez, A.K.; Franco, J.A.; Fernandez-Jaramillo, A.A.; Contreras-Medina, L.M.; Guevara-González, R.G.; Hernandez-Escobedo, Q. Machine learning for plant stress modeling: A perspective towards hormesis management. *Plants* **2022**, *11*, 970. [[CrossRef](#)] [[PubMed](#)]
12. Yoshioka, H.; Debeljak, P.; Prado, S.; Fuji, Y.; Ichihashi, Y.; Iwata, H. Interpretable multi-omics machine learning reveals drought-driven shifts in plant-microbe interactions. *Environ. Microbiome* **2026**, *21*, 57. [[CrossRef](#)] [[PubMed](#)]

13. Mimić, G.; Brdar, S.; Brkić, M.; Panić, M.; Marko, O.; Crnojević, V. Engineering meteorological features to select stress tolerant hybrids in maize. *Sci. Rep.* **2020**, *10*, 3421. [[CrossRef](#)] [[PubMed](#)]
14. Shetty, N.P.; Muniyal, B.; Sriyans, K.; Garg, K.; Pratap, S.; Priyanshu, A.; Kumar, D. Counterfactual based approaches for feature attributions of stress factors affecting rice yield. *Eng. Rep.* **2025**, *7*, e13085. [[CrossRef](#)]
15. Filippi, P.; Whelan, B.M.; Bishop, T.F.A. Explainable machine learning to map the impact of weather and soil on wheat yield and revenue across the eastern Australian grain belt. *Agriculture* **2024**, *14*, 2318. [[CrossRef](#)]
16. Li, S.; Li, Y.; Zhang, G.; Wang, L.; Song, M.; Fan, Y.; Siddique, K.H. Compound dry/wet and hot extremes decreased wheat/maize yield revealed by SHAP-RF and R-Vine Copula. *Field Crops Res.* **2025**, *334*, 110161. [[CrossRef](#)]
17. Wang, H.J.; Yu, Y.; Zhao, Y.; Wang, H.; Hu, X.; Ullah, U.; Yang, B.; Huang, J.; Chen, K.; Khan, A.H.; et al. Multilevel analysis of salt stress responses in sorghum during seed germination. *Front. Plant Sci.* **2025**, *16*, 1802398. [[CrossRef](#)] [[PubMed](#)]
18. Fei, L.; Bai, X.; Li, M.; Hao, A.; Xing, B. Decoding the saline-alkaline tolerance nexus in soybean: A dual-method evaluation model coupled with co-expression networks identifies core regulatory genes. *Front. Plant Sci.* **2025**, *16*, 1699904. [[CrossRef](#)] [[PubMed](#)]
19. Mannan, M.A.; Karim, M.A.; Haque, M.M.; Khaliq, Q.A.; Higuchi, H.; Nawata, E. Response of soybean to salinity: I. Genotypic variations in salt tolerance at the vegetative stage. *Trop. Agric. Dev.* **2012**, *56*, 117–122. [[CrossRef](#)]
20. Alzahrani, Y. Evaluation of drought and salinity tolerance potentials of different soybean genotypes based upon physiological, biochemical, and genetic indicators. *Front. Plant Sci.* **2024**, *15*, 1466363. [[CrossRef](#)] [[PubMed](#)]
21. Valencia, R.; Chen, P.; Ishibashi, T.; Conatser, M. A rapid and effective method for screening salt tolerance in soybean. *Crop Sci.* **2008**, *48*, 1773–1779. [[CrossRef](#)]
22. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*; Morgan Kaufmann: San Francisco, CA, USA, 1995; pp. 1137–1143.
23. Wang, B.; Meng, F.; Cheng, T.; Niu, J.; Rao, D.; Han, Z.; Zhang, W.; Zhang, Z. Comprehensive responses of physiology and rhizosphere microbiome to saline-alkaline stress in soybean seedlings with different tolerances. *Plants* **2025**, *14*, 3480. [[CrossRef](#)] [[PubMed](#)]
24. Nguyen, G.T.; Nguyen, Y.T.H.; Nguyen, H.D.; Chu, M.H.; Nguyen, Q.H. The DREB7 transcription factor enhances salt tolerance in soybean plants under salt stress. *Open Life Sci.* **2025**, *20*, 20251153. [[CrossRef](#)] [[PubMed](#)]
25. Xu, C.; Shan, J.; Liu, T.; Wang, Q.; Ji, Y.; Zhang, Y.; Wang, M.; Xia, N.; Zhao, L. CONSTANS-LIKE 1a positively regulates salt and drought tolerance in soybean. *Plant Physiol.* **2023**, *191*, 2427–2446. [[CrossRef](#)] [[PubMed](#)]
26. Yu, Z.; Niu, L.; Cai, Q.; Wei, J.; Shang, L.; Yang, X.; Ma, R. Improved salt-tolerance of transgenic soybean by stable over-expression of *AhBADH* gene from *Atriplex hortensis*. *Plant Cell Rep.* **2023**, *42*, 1291–1310. [[CrossRef](#)] [[PubMed](#)]
27. Dong, T.; Yan, L.; Wang, J.; Niu, Y.; Wang, L. Salt tolerance in soybean (*Glycine max* L.): A comprehensive review of molecular mechanisms, key regulators, and future perspectives for saline soil utilization. *Plants* **2025**, *14*, 3668. [[CrossRef](#)] [[PubMed](#)]
28. Sun, T.; Ma, N.; Wang, C.; Fan, H.; Wang, M.; Zhang, J.; Cao, J.; Wang, D. A Golgi-localized sodium/hydrogen exchanger positively regulates salt tolerance by maintaining higher K⁺/Na⁺ ratio in soybean. *Front. Plant Sci.* **2021**, *12*, 638340. [[CrossRef](#)] [[PubMed](#)]
29. Arif, M.; Khurshid, H.; Rabbani, M.A.; Jahanzaib, M.; Khattak, S.H.; Ullah, I.; Farah, M.A.; Singh, R.; Li, L. Morpho-physiological traits and stress indices reveal divergence in early-stage salt stress response in elite soybean germplasm. *J. King Saud Univ. Sci.* **2024**, *36*, 103522. [[CrossRef](#)]
30. Hoque, M.M.I.; Zheng, J.; Wang, G.Y. Evaluation of salinity tolerance in maize (*Zea mays* L.) genotypes at seedling stage. *J. BioSci. Biotechnol.* **2015**, *4*, 39–49. [[CrossRef](#)]
31. Mahboob, W.; Khan, M.A.; Shirazi, M.U.; Mumtaz, S.; Shereen, A. Using growth and ionic contents of wheat seedlings as rapid screening tool for salt tolerance. *J. Crop Sci. Biotechnol.* **2018**, *21*, 173–181. [[CrossRef](#)]
32. Kakar, N.; Jumaa, S.H.; Redoña, E.D.; Warburton, M.L.; Reddy, K.R. Evaluating rice for salinity using pot-culture provides a systematic tolerance assessment at the seedling stage. *Rice* **2019**, *12*, 57. [[CrossRef](#)] [[PubMed](#)]
33. Johnson, K.; Vu, D.H.; Asch, F. Traits contributing to salinity tolerance in rice genotypes from the Mekong Delta. *J. Agron. Crop Sci.* **2024**, *210*, e12679. [[CrossRef](#)]
34. Kiani-Pouya, A.; Rasouli, F. The potential of leaf chlorophyll content to screen bread-wheat genotypes in saline condition. *Photosynthetica* **2014**, *52*, 288–300. [[CrossRef](#)]
35. Kou, X.; Han, W.; Kang, J. Responses of root system architecture to water stress at multiple levels: A meta-analysis of trials under controlled conditions. *Front. Plant Sci.* **2022**, *13*, 1085409. [[CrossRef](#)] [[PubMed](#)]
36. Bargali, K.; Tewari, A. Growth and water relation parameters in drought-stressed *Coriaria nepalensis* seedlings. *J. Arid Environ.* **2004**, *58*, 505–512. [[CrossRef](#)]
37. Asekova, S.; Kulkarni, K.P.; Patil, G.; Kim, M.; Song, J.T.; Nguyen, H.T.; Shannon, J.G.; Lee, J.D. Genetic analysis of shoot fresh weight in a cross of wild (*G. soja*) and cultivated (*G. max*) soybean. *Mol. Breed.* **2016**, *36*, 103. [[CrossRef](#)]
38. Haq, T.U.; Akhtar, J.; Gorham, J.; Steele, K.A.; Khalid, M. Genetic mapping of QTLs controlling shoot fresh and dry weight under salt stress in rice (*Oryza sativa* L.) cross between Co39 × Moroberekan. *Pak. J. Bot.* **2008**, *40*, 2369–2381.

39. Khan, M.M.; Rahman, M.M.; Hasan, M.M.; Amin, M.F.; Matin, M.Q.I.; Faruq, G.; Alkeridis, L.A.; Gaber, A.; Hossain, A. Assessment of the salt tolerance of diverse bread wheat (*Triticum aestivum* L.) genotypes during the early growth stage under hydroponic culture conditions. *Heliyon* **2024**, *10*, e29042. [[CrossRef](#)] [[PubMed](#)]
40. Majeed, A.; Anwar-ul-Haq, M.; Akhtar, J.; Basra, S.M.A. Screening of maize genotypes against poor quality water in solution culture. *J. Agric. Res.* **2014**, *52*, 357–368.
41. Yan, C.; Song, S.; Wang, W.; Wang, C.; Li, H.; Wang, F.; Li, S.; Sun, X. Screening diverse soybean genotypes for drought tolerance by membership function value based on multiple traits and drought-tolerant coefficient of yield. *BMC Plant Biol.* **2020**, *20*, 321. [[CrossRef](#)] [[PubMed](#)]
42. Suriya-Arunroj, D.; Supapoj, N.; Vanavichit, A.; Toojinda, T. Screening and selection for physiological characters contributing to salinity tolerance in rice. *Kasetsart J. Nat. Sci.* **2005**, *39*, 174–185.
43. Nampei, M.; Ueda, A. Dataset of rice growth for saline-alkaline tolerance screening. *Data Brief.* **2025**, *62*, 111998. [[CrossRef](#)] [[PubMed](#)]
44. Munns, R.; Passioura, J.B.; Colmer, T.D.; Byrt, C.S. Osmotic adjustment and energy limitations to plant growth in saline soil. *New Phytol.* **2020**, *225*, 1091–1096. [[CrossRef](#)] [[PubMed](#)]
45. Shelke, D.B.; Chambhare, M.R.; Nikalje, G.C. *Wild and Cultivated Soybean: Mechanisms of Salt Tolerance and Agricultural Applications*; IntechOpen: London, UK, 2026.
46. Yuan, Y.; Zhu, L.; Cao, B.; You, J.; Zhi, H.; Li, K.; Gu, W.; Yang, H.; Li, C. Integrated multi-omics analysis revealed the synergistic regulatory mechanisms of salt tolerance in soybean (Kefeng 1). *Plants* **2026**, *15*, 555. [[CrossRef](#)] [[PubMed](#)]
47. Meng, Y.; Xin, N.; Li, Y.; Zhang, Y.; Zhou, P.; Jia, H.; Xiong, S.; Luan, H.; Zhao, L.; Wu, G. Salt tolerance and oil quality stability of gene-edited high-oleic soybean lines under saline conditions. *Field Crops Res.* **2026**, *343*, 110498. [[CrossRef](#)]
48. Yu, Z.; Wei, J.; Cai, Q.; Zhang, Y.; Zhao, Q.; Niu, L.; Ren, P.; Shang, L.; Yang, X.; Dong, Y.; et al. SchAL1-mediated enhancement of salt tolerance in soybean: From stable transgenic inheritance to field trial validation. *Plant Biotechnol. J.* **2026**, *24*, 1750–1765. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.