

Article

Explainable Machine Learning Prediction of Soybean Lodging Grade and Key Trait Analysis Under High-Density Drip Irrigation Cultivation

Xiangchi Zhang, Xin Su, Hengbin Zhang, Jing Zhao, You Ge, Zhanqin Zhang, Kai Zeng and Yong Zhan *

Xinjiang Academy of Agricultural and Reclamation Science, Shihezi 832000, China; 17699537028@163.com (X.Z.); 18209933431@163.com (X.S.); shzzhbb@163.com (H.Z.); zhaojing_yu0818@163.com (J.Z.); 18119157351@163.com (Y.G.); zzq3000qwe@163.com (Z.Z.); 15299949856@163.com (K.Z.)

* Correspondence: shzzhy@163.com

Abstract

To establish an accurate and interpretable prediction framework for soybean lodging grade and clarify the core regulatory traits and differentiated driving mechanisms of soybean lodging under high-density drip irrigation cultivation, 356 spring soybean germplasm accessions were used as experimental materials in this study. Morphological and mechanical traits including plant height (PH), stem pulling force (SPF), internode number (IN) and petiole length (PL) were measured over two consecutive years of field phenotyping. Two composite evaluation indices, plant height/stem pulling force ratio (PH/SPF) and plant height/internode number ratio (PH/IN), were further constructed. Four machine learning algorithms were adopted to develop multi-classification models for soybean lodging grade prediction. SHAP analysis combined with three global sensitivity approaches (perturbation analysis, Sobol' method and Morris screening) was applied to decipher the regulatory patterns of key traits. The results showed that lodging grade significantly affected soybean grain yield and explained 25–28% of the phenotypic yield variation; yield reduction tended to plateau under severe lodging. Compared with single indicators such as SPF and PL, the two derived composite indices could stably distinguish soybean accessions with different lodging grades and exhibited stronger discriminatory power. Model comparison revealed that the XGBoost model achieved optimal prediction accuracy and generalization stability for lodging grade, with a weighted F1-score of 95.34% on the test set, significantly outperforming the conventional linear model. Interpretability analysis demonstrated that the PH/IN, PH, and PH/SPF acted as the primary positive traits promoting lodging, while SPF was the sole protective trait. Driving factors of lodging presented obvious gradient heterogeneity: mild lodging was dominated by the imbalance of plant architecture ratio, whereas severe lodging was governed by the cumulative effects of PH and IN. Strong interactions existed among all measured traits. The interpretable machine learning framework established in this study can provide theoretical support and technical references for lodging-resistant germplasm screening and targeted plant architecture regulation for densely planted soybean under drip irrigation systems.



Academic Editor: Chengming Sun

Received: 24 July 2026

Revised: 8 August 2026

Accepted: 21 August 2026

Published: 26 August 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 \(CC BY\)](#) license.

Keywords: soybean (*Glycine max* (L.) Merr.); lodging grade; composite trait; machine learning; XGBoost

1. Introduction

Soybean (*Glycine max* (L.) Merr.) is a globally critical grain and oilseed crop that occupies a central position in agricultural production and global food security. It supplies a substantial proportion of dietary plant protein for human populations and constitutes a fundamental feedstock for the vegetable oil and animal feed sectors [1]. Nonetheless, lodging remains a predominant yield-limiting factor in intensive high-yield soybean production systems. Accumulating evidence indicates that soybean lodging arises from the deterioration of stem mechanical resistance driven by alterations in plant height, internode number, stem structural strength and ambient growth conditions [2]. Lodging events occurring between the pod-setting and seed-filling stages can reduce soybean yield by up to 50% or more [3,4]. Conversely, reduced petiole length contributes to a more compact canopy architecture and improves lodging tolerance under high-density cropping systems [5]. In high-density super-high-yield production systems with mulched drip irrigation, which feature high fertilizer application rates and dense planting patterns, soybean lodging poses an especially prominent challenge. Exuberant vegetative growth and densely closed crop canopies substantially exacerbate lodging risk, rendering it a critical bottleneck limiting the stability and continued enhancement of soybean productivity [6]. Accordingly, establishing accurate lodging risk prediction and targeted prevention strategies has become an urgent research priority for soybean breeding and precision agronomic management under such cultivation systems.

Conventional investigations into soybean lodging resistance have predominantly relied on field phenotypic evaluation. One mainstream approach assesses lodging potential by correlating lodging grades with morphological and biomass traits, including plant height, above- and below-ground fresh and dry weights, tissue water content, and stem mechanical strength [7–9]. A second line of research evaluates lodging risk from a mechanistic perspective, using stem biochemical and anatomical traits such as lignin and cellulose concentrations, activities of key rate-limiting enzymes involved in cell wall biosynthesis, and stem anatomical features [10–12]. Nevertheless, both analytical frameworks depend on destructive sampling protocols, which prevent repeated phenotyping of individual plants and impede continuous, longitudinal tracking of lodging development over the growth cycle.

From a modeling perspective, lodging arises from the coupling of three-dimensional plant structural architecture and dynamic external mechanical forces. Most current morphogenetic models adopt two-dimensional simplifications, which limits their ability to fully recapitulate complex lodging behaviors including stem buckling, canopy structural rearrangement, and root anchorage failure [13]. However, three-dimensional reconstruction techniques suffer from inherent technical limitations, notably the trade-off between spatial resolution and imaging coverage, as well as insufficient capability to capture dynamic lodging processes, which further restricts the predictive accuracy of process-based lodging models [14]. Luo et al. [15] further noted that lodging resistance evaluation models derived from conventional statistical methods [16]—including membership function analysis, discriminant function analysis and logistic regression—are predominantly parameterized using static datasets under ideal experimental conditions, and thus fail to deliver accurate predictions to support real-time field lodging management. While correlation-based agronomic trait analysis can elucidate inter-trait associations and guide breeding selection, it is inherently limited in resolving nonlinear relationships among multiple variables and cannot effectively disentangle complex trait interactions, representing a fundamental constraint on prediction precision [4,17].

Advances in information technology have positioned machine learning as a powerful data mining and pattern recognition tool that has demonstrated substantial application

potential across multiple agricultural domains [18]. Machine learning algorithms can uncover underlying patterns from complex nonlinear datasets and construct robust predictive models, offering novel analytical pathways for addressing intricate challenges in agricultural science [19]. Mourtzinis et al. demonstrated that ensemble learning algorithms excel at capturing nonlinear relationships and complex interactive effects among agricultural variables [20]. Wang et al. [21] employed a particle swarm optimization-back propagation neural network to quantify the correlation between wind-induced response traits and lodging index in rapeseed seedlings, achieving superior predictive performance for lodging index estimation. In recent years, remote sensing technology has emerged as a promising approach for graded monitoring of crop lodging disasters [22]. Sarkar et al. [4] developed a lodging scoring model with 96% classification accuracy using unmanned aerial vehicle imagery captured during the reproductive stage and an artificial neural network algorithm in soybean breeding fields. However, studies on rice lodging prediction have indicated that conventional image analysis approaches suffer from limited real-time performance and high equipment costs, and thus cannot support rapid, convenient and low-cost lodging risk assessment at the field scale [23].

Despite these advances, two critical research gaps remain in current soybean lodging prediction studies. On the one hand, most evaluations of lodging potential are still based on isolated morphological or mechanical traits alone [7]; composite ratio indices integrating plant architectural attributes and stem mechanical properties have rarely been systematically constructed and validated, which constrains the comprehensive accuracy and stability of lodging risk assessment. On the other hand, relevant investigations focusing on soybean lodging prediction in mulched drip-irrigated high-density systems are still very limited, and no specialized prediction framework has been established for this cultivation pattern. Accordingly, to address the aforementioned technical limitations while balancing prediction accuracy, field operability and implementation cost, this study selected four easily measurable representative traits: plant height, stem pulling force, petiole length, and internode number, and constructed an integrated quantitative phenotypic prediction system combining morphological traits, mechanical properties and derived composite indicators. Three ensemble learning models (Random forest, XGBoost, LightGBM) and a conventional logistic regression model were employed to elucidate the response patterns between multi-dimensional traits and lodging risk grades, identify the optimal lodging prediction framework, and screen key regulatory features. The findings provide a scientific reference for breeding lodging-resistant soybean varieties and optimizing agronomic practices for population lodging prevention in high-density mulched drip irrigation systems.

2. Materials and Methods

2.1. Experimental Materials and Design

The experimental materials consisted of 356 spring soybean varieties (lines) suitable for cultivation under oasis irrigated ecological conditions, and detailed germplasm information is provided in Table A1. Two consecutive years of field trials were carried out from 2024 to 2025 at the Shihezi Experimental Base (E 85.9800, N 44.3160). The experiment adopted a randomized block design with manual hill seeding. Each plot contained six rows planted in a wide-narrow row patterns. The plant spacing was 7.5 cm, the average row spacing was 38.3 cm, and the theoretical seedling retention density was 345,000 plants/hm². Three biological replicates were set for each germplasm (Figure 1). Sowing was conducted on 1 May, and harvesting was completed in mid-October.

Field management throughout the growing season followed the technical regulations for integrated water and fertilizer management under mulched drip irrigation. Two chemical regulation treatments were implemented at the initial flowering stage and 3 d

before the initial podding stage, with mepiquat chloride applied as a foliar spray at rates of 150 g/hm² and 300 g/hm², respectively. Fertilizers were applied with irrigation water 12 times during the growing season, with cumulative application rates of 660 kg/hm² urea, 555 kg/hm² monoammonium phosphate, and 570 kg/hm² potassium sulfate.

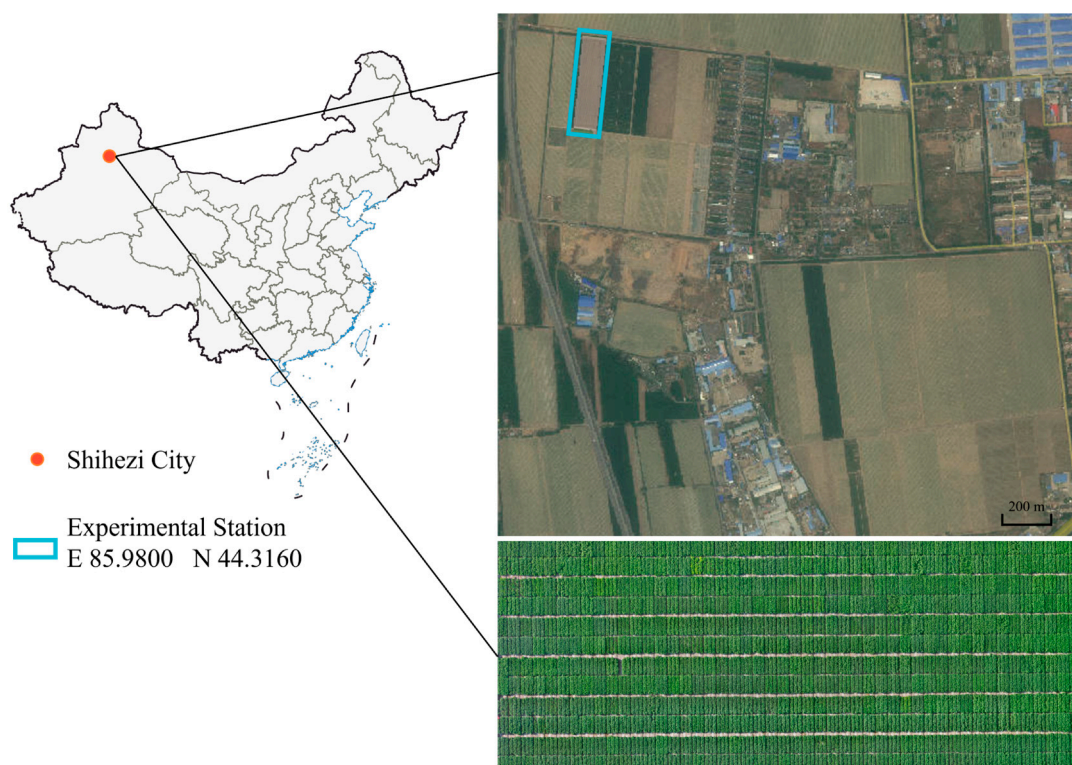


Figure 1. Experimental site information and field layout.

2.2. Measurement Indicators and Methods

Lodging-related morphological indicators and stem mechanical indicators of soybean plants were measured at the early seed-filling stage; at the maturity stage, the population lodging grade of each test accession was assessed and plot yield was determined.

2.2.1. Lodging Grade

Plants were defined as lodged when the inclination angle between the main stem and the ground was less than 30°. The percentage of lodged plants relative to the total number of plants in each plot was calculated. Lodging grades were classified based on the mean lodging percentage as follows: Grade 1 for 0 < lodging percentage < 25%; Grade 2 for 25% ≤ lodging percentage < 50%; Grade 3 for 50% ≤ lodging percentage < 75%; Grade 4 for lodging percentage ≥ 75% [24].

2.2.2. Morphological and Mechanical Traits

Plant height: Uniformly growing plants were selected, and the vertical distance from the cotyledonary node to the apical growing point of the plant was measured.

Stem pulling force: An internode at the middle section of the stem was selected and pulled horizontally at a constant speed using a tensile tester (YYD-1, Zhejiang Top Cloud-Agri Technology Co., Ltd., Hangzhou, China). The force required to produce a 75° stem bending deflection was recorded.

Petiole length: The third leaf from the top of the plant was selected, and the length from the petiole base to the leaf insertion site was measured.

Number of internode nodes: The total number of internodes from the cotyledonary node to the apical growing point of the stem was counted.

2.2.3. Yield

At full maturity, soybean grains were harvested separately for each plot and naturally air-dried until seed moisture content stabilized. The grain yield of each plot was weighed, averaged across replicates, and subsequently converted to yield per unit area (kg/hm^2).

2.3. Construction and Analysis of Lodging Risk Prediction Models

To address the inherent characteristics of soybean field phenotypic data, including strong nonlinearity among traits, imbalanced distribution of lodging grades, and inevitable measurement noise in field experiments, a customized stage analytical framework was constructed in this study rather than directly applying a generic machine learning workflow. As illustrated in Figure 2, the whole pipeline proceeds sequentially from data collection of core phenotypic indicators, through preliminary statistical analysis with Welch-corrected ANOVA, Bayesian ANOVA and Spearman correlation to explore inter-group differences and inter-trait associations, to trait preprocessing comprising isolation forest-based outlier removal, derived composite feature construction and Z-score standardization followed by stratified 8:2 train-test splitting. This is followed by multi-model comparative evaluation where four representative algorithms are built under a unified hyperparameter baseline and comprehensively assessed via multiple classification metrics, 10-fold stratified cross-validation and paired *t*-tests to identify the optimal prediction model, and finally key trait selection integrating the SHAP interpretability framework and three global sensitivity methods to cross-validate core regulatory traits and their interaction patterns and decipher the driving mechanism of soybean lodging.

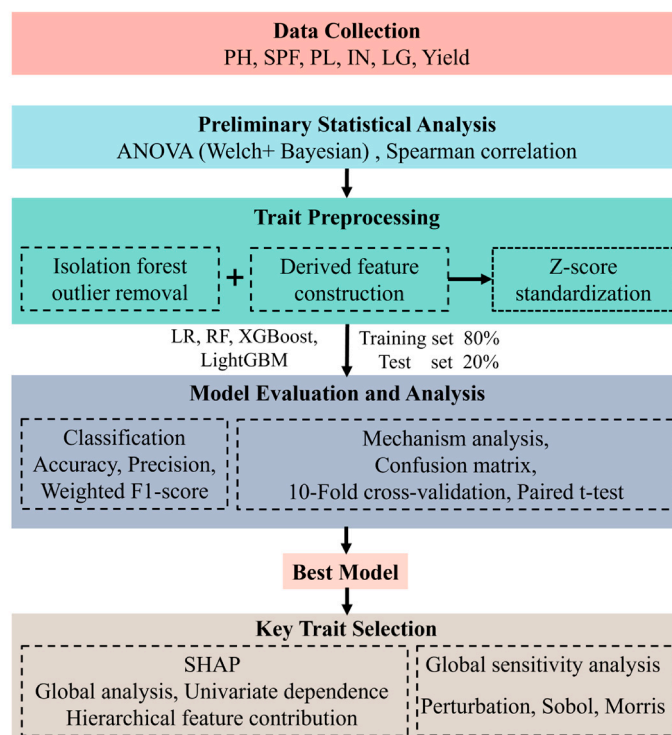


Figure 2. Schematic diagram of the research workflow. PH: plant height; SPF: stem pulling force; PL: petiole length; IN: internode number; LG: lodging grade; LR: logistic regression; RF: random forest; XGBoost: extreme gradient boosting; LightGBM: light gradient boosting machine.

Following the analytical workflow depicted in Figure 2, a sequential three-stage analytical pipeline was implemented, encompassing data preprocessing, comparative model construction and performance evaluation, and in-depth feature interpretation.

2.3.1. Data Preprocessing

The isolation forest algorithm was employed to detect and remove outlier samples from the feature set. Two derived ratio features, namely the plant height-to-stem pulling force ratio and the plant height-to-main stem internode number ratio, were calculated from the four original morphological and mechanical traits. A lodging grade prediction dataset was subsequently constructed by integrating the six features with the corresponding lodging grade labels.

To eliminate the influence of dimensional differences among features, all six features were standardized using the Z-score method:

$$Z = \frac{X - \mu}{\sigma}$$

where X represents the raw observed value of a given feature, μ denotes the mean value of the feature, and σ is the standard deviation of the feature.

The standardized dataset was partitioned into a training set and a test set at an 8:2 ratio via stratified sampling. This partitioning strategy maintained a consistent distribution of lodging grade samples across both sets, thereby mitigating prediction bias caused by class imbalance.

2.3.2. Model Construction

Four representative algorithms—logistic regression, random forest, extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM)—were selected to develop multi-class prediction models for lodging grades of spring soybean. The baseline hyperparameter configurations for each model are listed in Table 1.

Table 1. Hyperparameter configuration of each model.

Model Name	Hyperparameter Settings
Logistic regression	lbfgs solver, L2 penalty ($C = 1.0$), multinomial, max_iter = 1000, balanced class weights, random_state = 42
Random forest	n_estimators = 1200, max_depth = 12, min_samples_split = 5, max_features = 'sqrt', bootstrap = True, oob_score = True, subsample = 0.9, balanced class weights, random_state = 42
XGBoost	n_estimators = 1200, max_depth = 12, consistent sampling & class weight strategy as random forest, random_state = 42
LightGBM	n_estimators = 1300, learning_rate = 0.015, num_leaves = 40, max_depth = 7, subsample = 0.8, colsample_bytree = 0.7, reg_alpha = 1.0, reg_lambda = 0.8, min_child_samples = 25, balanced class weights, random_state = 42

2.3.3. Model Evaluation and Analysis

Three quantitative evaluation metrics, including overall accuracy, weighted precision, and weighted F1-score, were adopted to comprehensively assess the predictive performance of each multi-class classification model.

$$\text{Accuracy} = \frac{\sum_{i=1}^k TP_i}{N}$$

$$\text{Weighted precision} = \sum_{i=1}^k \left(\frac{TP_i}{TP_i + FP_i} \times \frac{N_i}{N} \right)$$

$$\text{Weighted F1 - score} = \sum_{i=1}^k \omega_i \times \frac{2 \times P_i \times R_i}{P_i + R_i}$$

where k is the total number of lodging grades; N is the total number of samples; and N_i is the number of true samples belonging to class i .

TP_i = true positives of class i (samples truly belonging to class i and predicted as class i);

FP_i = false positives of class i ;

$P_i = TP_i / (TP_i + FP_i)$ denotes precision for class i ;

$R_i = TP_i / (TP_i + FN_i)$ denotes recall for class i ;

FN_i = false negatives of class i ;

$\omega_i = N_i / N$ represents the sample weight of class i .

To evaluate the generalization stability of each candidate model and test the statistical significance of performance differences across the four classification models, 10-fold stratified cross-validation followed by pairwise paired t -tests was performed based on overall accuracy and weighted F1-score, with a significance level set at $p < 0.05$.

2.4. Feature Interpretation of the Optimal Model

The Shapley Additive exPlanations (SHAP) framework was employed to perform interpretability analysis of the optimal prediction model and elucidate the mechanistic contributions of individual agronomic traits to lodging grade prediction [25]. Three complementary sensitivity analysis methods were further integrated to quantify the impacts of feature perturbations on model outputs, mutually validate the contribution patterns of input traits, and verify the robustness of model predictions [26].

Perturbation Analysis: The mean value of each trait in the training set was set as the baseline. A $\pm 10\%$ perturbation was applied to each single trait independently, and the magnitude of variation in the model prediction results was recorded. Normalized sensitivity scores were subsequently calculated to rank all traits by their sensitivity level.

Sobol Variance Decomposition Method: A total of 512 sample sets were generated using the Saltelli sampling scheme to decompose the variance of model outputs. Two sensitivity indices were estimated: the first-order sensitivity index S_1 , which quantifies the independent contribution of each trait, and the total sensitivity index S_T , which captures the overall contribution incorporating interaction effects. This decomposition enables explicit differentiation between the independent effects and interactive effects of input traits.

Morris Screening Method: The analysis was configured with 10 trajectories and 4-level discrete sampling. Two core metrics were derived: the mean absolute elementary effect μ^* , which characterizes the overall sensitivity of each trait, and the standard deviation of elementary effects σ , which reflects the strength of nonlinearity and interaction effects among traits. This method supports efficient and rapid identification of core sensitive traits from the feature set.

The combination of SHAP analysis and three complementary sensitivity methods provides complementary insights: perturbation analysis intuitively reflects the fluctuation of prediction results caused by trait changes, Sobol decomposition quantifies the independent and interactive contributions of each trait separately, and Morris screening efficiently identifies core sensitive features with low computational cost. The consistent conclusions obtained from multiple methodological perspectives effectively avoid the one-sidedness of a single interpretability method and ensure the reliability of the identified key lodging regulatory traits.

2.5. Data Processing and Statistical Analysis

Basic data collation was performed using Microsoft Excel 2011. Given the uneven sample distribution across lodging grades and unequal variances among groups, Welch-corrected one-way analysis of variance (ANOVA) followed by Games–Howell post hoc multiple comparisons was conducted using SPSS 22.0 to compare soybean morphological traits, mechanical traits, derived indicators, and yield among different lodging grade groups. Bayesian ANOVA was implemented in R 4.3.1 with a zero-scaled Cauchy prior set at $r_{\text{Cauchy}}^{\text{JZS}} = 0.71$. The natural Bayes factor $\log_e(\text{BF}_{01})$ and the posterior Bayesian coefficient of determination $\hat{R}_{\text{Bayesian}}^{\text{posterior}}$ were calculated to cross-validate the conclusions regarding inter-group differences. Machine learning model construction, performance evaluation, and feature interpretation of the optimal model were performed using Python 3.9 deployed on the Anaconda platform (version 2.5.2). All data visualization plots were generated using Microsoft Excel 2011, Python 3.9, and the Hiplot web platform (<https://hiplot.com.cn/>).

3. Results

3.1. Annual Distribution of Lodging Grades and Yield Response Patterns in Soybean Germplasm

Under the high-density cultivation regime with integrated water and fertigation management via mulched drip irrigation, spring soybean canopies develop high closure and face pronounced lodging risk. As illustrated in Figure 3a, the overall germplasm population across the two years, 2024 and 2025, exhibited a consistent pattern: the number of accessions decreased with increasing lodging grade. Nevertheless, distinct inter-annual variation was observed in the composition of lodging grades. In 2024, the number of accessions assigned to lodging Grades 1–4 was 197, 84, 59, and 16, respectively. The germplasm panel was dominated by accessions with mild lodging, with a coefficient of variation (CV) of 86.83% for lodging grade, indicating high phenotypic dispersion of lodging performance among the tested materials. In 2025, the corresponding counts for Grades 1–4 were 136, 97, 77, and 46, respectively. The overall lodging severity of the population increased in this year, while the CV declined to 42.37%, reflecting a more concentrated distribution of lodging grades across the panel.

Lodging severity exerted a significant effect on soybean yield formation (Figure 3b). Consistent patterns were observed across the two experimental years: soybean yield declined continuously with ascending lodging grades. In 2024, accessions with Grade 1 lodging exhibited the highest yield, and yield differences between all adjacent lodging grades reached an extremely significant level ($p < 0.01$). The yield trend in 2025 was aligned with that in 2024, whereas no statistically significant yield difference was detected between Grade 3 and Grade 4 lodging accessions. This indicates that once lodging progresses into the severe range, further aggravation of lodging leads to a diminished rate of yield decline. The partial omega-squared effect sizes $\hat{\omega}_p^2$ for both years exceeded 0.14, reaching the threshold for a strong effect and confirming that lodging grade is a dominant factor driving yield differentiation among germplasm accessions. For both datasets, the natural Bayes factors $\log_e(\text{BF}_{01})$ were below zero, and the posterior Bayesian coefficients of determination $\hat{R}_{\text{Bayesian}}^{\text{posterior}}$ were 0.25 and 0.28, respectively. These results corroborate the presence of robust yield differences among lodging grades. Lodging grade accounted for 25–28% of the phenotypic variation in yield, demonstrating a stable quantitative association between lodging severity and yield performance.

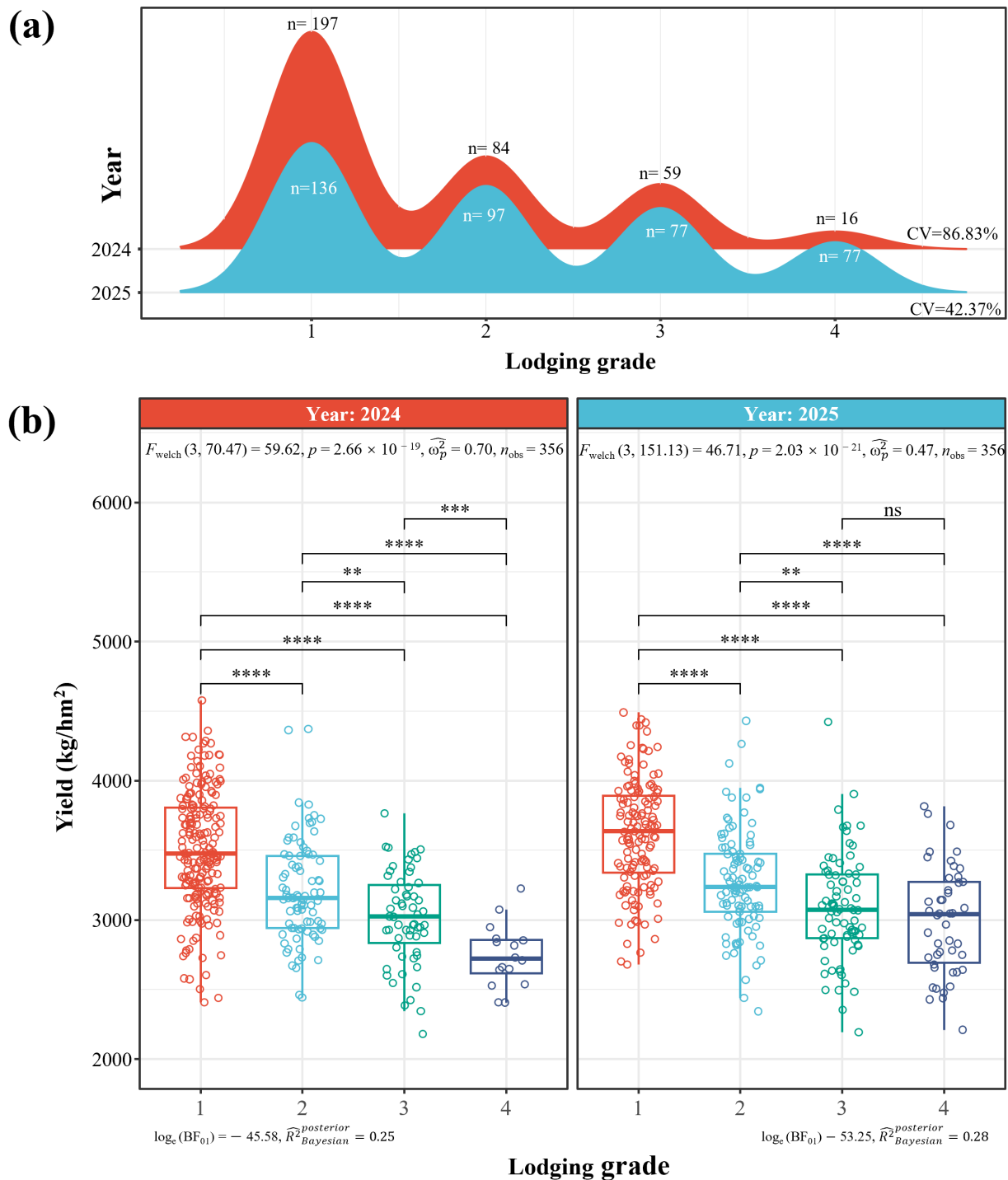


Figure 3. Distribution characteristics of lodging grades in soybean germplasm across two experimental years (a) and yield differences among different lodging grades (b). Phenotypic evaluation of lodging traits and yield was performed on 356 soybean germplasm accessions under a high-density ultra-high-yield cultivation regime with mulched drip irrigation from 2024 to 2025. Lodging grades were classified as follows: Grade 1, proportion of lodged plants < 25%; Grade 2, proportion of lodged plants 25–50%; Grade 3, proportion of lodged plants 50–75%; Grade 4, proportion of lodged plants ≥ 75%. Yield differences among lodging grades were analyzed using Welch-corrected one-way ANOVA, Games–Howell post hoc multiple comparisons, and Bayesian ANOVA. Significance levels between groups are indicated by connecting lines; ns denotes no significant difference, while **, ***, and **** represent $p < 0.01$, $p < 0.001$, and $p < 0.0001$, respectively. CV, coefficient of variation; $\hat{\omega}_p^2$, partial omega-squared effect size; $\log_e(BF_{01})$, natural log-transformed Bayes factor; $\hat{R}_{Bayesian}^{2\text{posterior}}$, posterior Bayesian coefficient of determination.

3.2. Phenotypic Differentiation of Agronomic and Stem Mechanical Traits in Soybean Germplasm with Different Lodging Grades

Across the two-year field trials under the mulched drip irrigation cultivation regime, soybean germplasm accessions assigned to different lodging grade groups exhibited distinct phenotypic differentiation in agronomic traits and stem mechanical indicators (Figure 4).

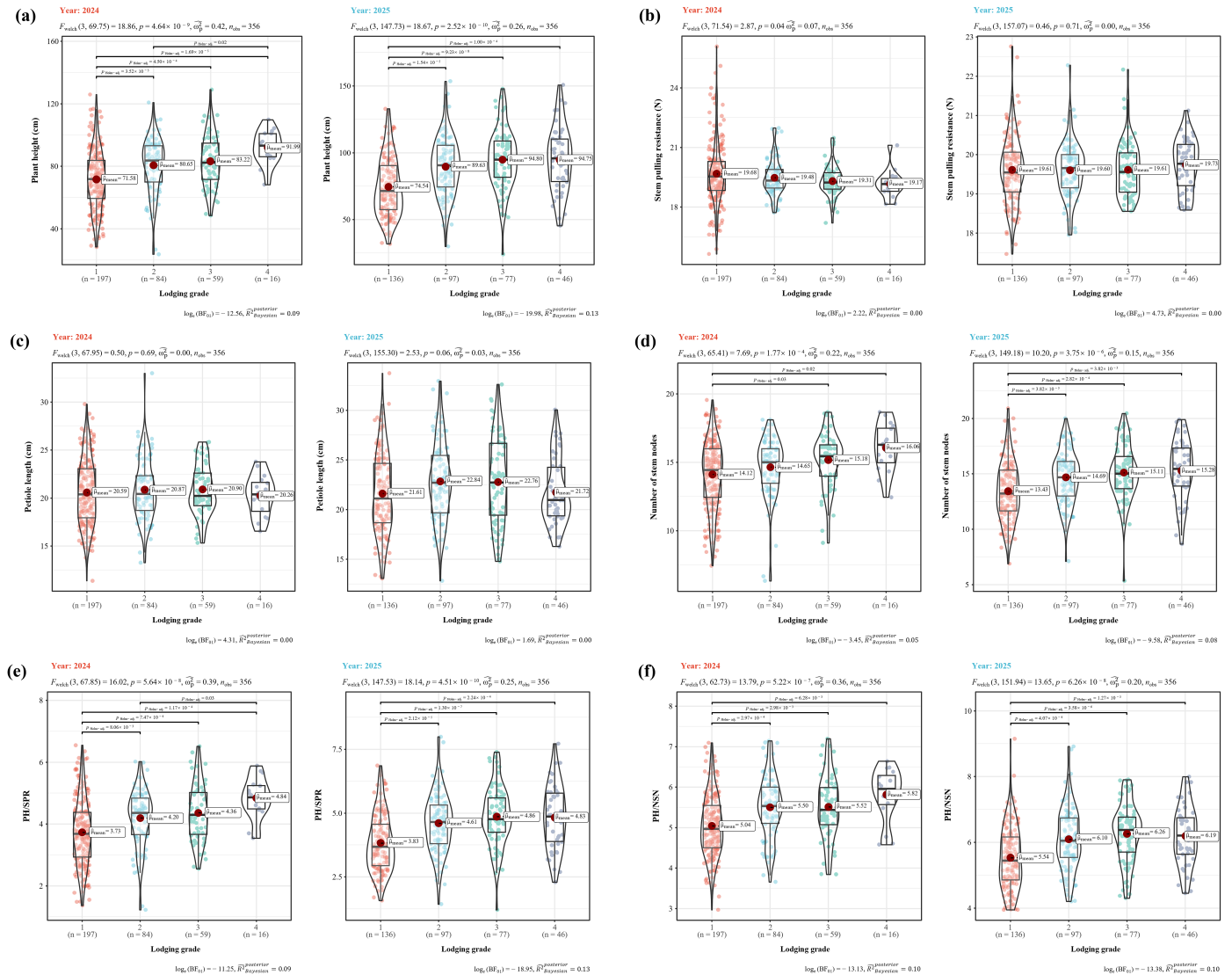


Figure 4. Phenotypic distribution and inter-group differences in morphological and stem mechanical traits in soybean across different lodging grades. Phenotypic evaluation of plant height (a), stem pulling force (b), petiole length of the third leaf from the top (c), internode number (d), PH/SPF (e) and PH/IN (f) was conducted on 356 soybean germplasm accessions under a high-density ultra-high-yield cultivation regime with mulched drip irrigation from 2024 to 2025. Lodging grades were classified by the following criteria: Grade 1, proportion of lodged plants < 25%; Grade 2, proportion of lodged plants 25–50%; Grade 3, proportion of lodged plants 50–75%; Grade 4, proportion of lodged plants $\geq$ 75%. $\hat{\omega}_p^2$, partial omega-squared effect size; $\log_e(\text{BF}_{01})$, natural log-transformed Bayes factor; $\hat{R}_{\text{Bayesian}}^2$, posterior Bayesian coefficient of determination; PH, plant height; SPF, stem pulling force; SL, petiole length; IN, number of main stem nodes; PH/SPF, plant height-to-stem pulling force ratio; PH/IN, plant height-to-internode number ratio.

Plant height, a core determinant of aboveground center of gravity and lodging risk, increased progressively with lodging grade. In 2024, mean plant height ranged from 71.58 to 91.99 cm for Grades 1–4, with significant differences between most groups ($p < 0.05$). In 2025, overall plant height was higher, and Grade 1 differed extremely significantly from all other grades ($p < 0.0001$). Partial omega-squared values $\hat{\omega}_p^2$ were 0.42 and 0.36 for the two years (both strong effects), supported by negative Bayes factors $\log_e$ (BF_{01}) and $\hat{R}_{Bayesian}^{2,posterior}$ of 0.09 and 0.13, confirming greater plant height in higher-lodging-grade germplasm (Figure 4a).

Stem pulling force, a direct measure of stem mechanical strength, showed limited inter-group differentiation. A weak significant difference was observed in 2024 ($p = 0.04$) but not in 2025 ($p = 0.72$), with $\hat{\omega}_p^2$ values of 0.00–0.07 (negligible effects) and positive $\log_e$ (BF_{01}) values. Bayesian analysis did not support substantial inter-group differences, indicating that stem pulling force alone cannot reliably distinguish lodging potential (Figure 4b).

Internode number, linked to plant structural stability, showed strong effect sizes ($\hat{\omega}_p^2 = 0.22, 0.15$) across years (Figure 4c).

Higher lodging grades generally had more nodes, with Grade 1 differing significantly from higher grades (Figure 4d).

Two derived composite indicators (PH/SPF, PH/IN) consistently showed strong effects ($\hat{\omega}_p^2 = 0.20$ – 0.39), with Grade 1 having significantly lower values than all other groups ($p < 0.001$). They outperformed single stem pulling force measurements in differentiating lodging potential (Figure 4e,f).

Collectively, plant height and its composite indicators were stably associated with lodging risk across years, whereas stem pulling force and petiole length exhibited strong inter-annual variability. Soybean lodging arises from the combined effects of morphological traits, stem mechanical properties and environmental variation.

3.3. Correlation Analysis of Agronomic Traits, Lodging Grade and Yield

A pairwise Spearman correlation matrix was constructed to elucidate interrelationships among morphological traits, stem mechanical traits, derived composite indicators, lodging grade and grain yield. Plant height, internode number and grain yield followed an approximately normal distribution, whereas stem pulling force, petiole length of the uppermost third leaf and PH/SPF were right-skewed (Figure 5). Lodging grade, a discrete ordinal variable, was dominated by mild lodging samples, matching typical field occurrence patterns.

PH/SPF showed the strongest positive correlation with lodging grade ($\rho = 0.737$), markedly outperforming single stem pulling force ($\rho = -0.060$). Strong positive correlations between plant height and both composite indicators suggested potential multicollinearity risk. Lodging grade was significantly negatively correlated with grain yield ($\rho = -0.435$), with petiole length contributing minimally to lodging grade.

3.4. Performance Comparison and Evaluation of Multiple Models for Soybean Lodging Grade Prediction

Using six features as inputs, a multi-class prediction framework for soybean lodging grade was established, incorporating one linear model (logistic regression) and three ensemble tree-based models: random forest, XGBoost and LightGBM. Stratified 10-fold cross-validation was employed, and the prediction performance of each model was compared across three evaluation metrics: accuracy, weighted precision and weighted F1-score (Figure 6).

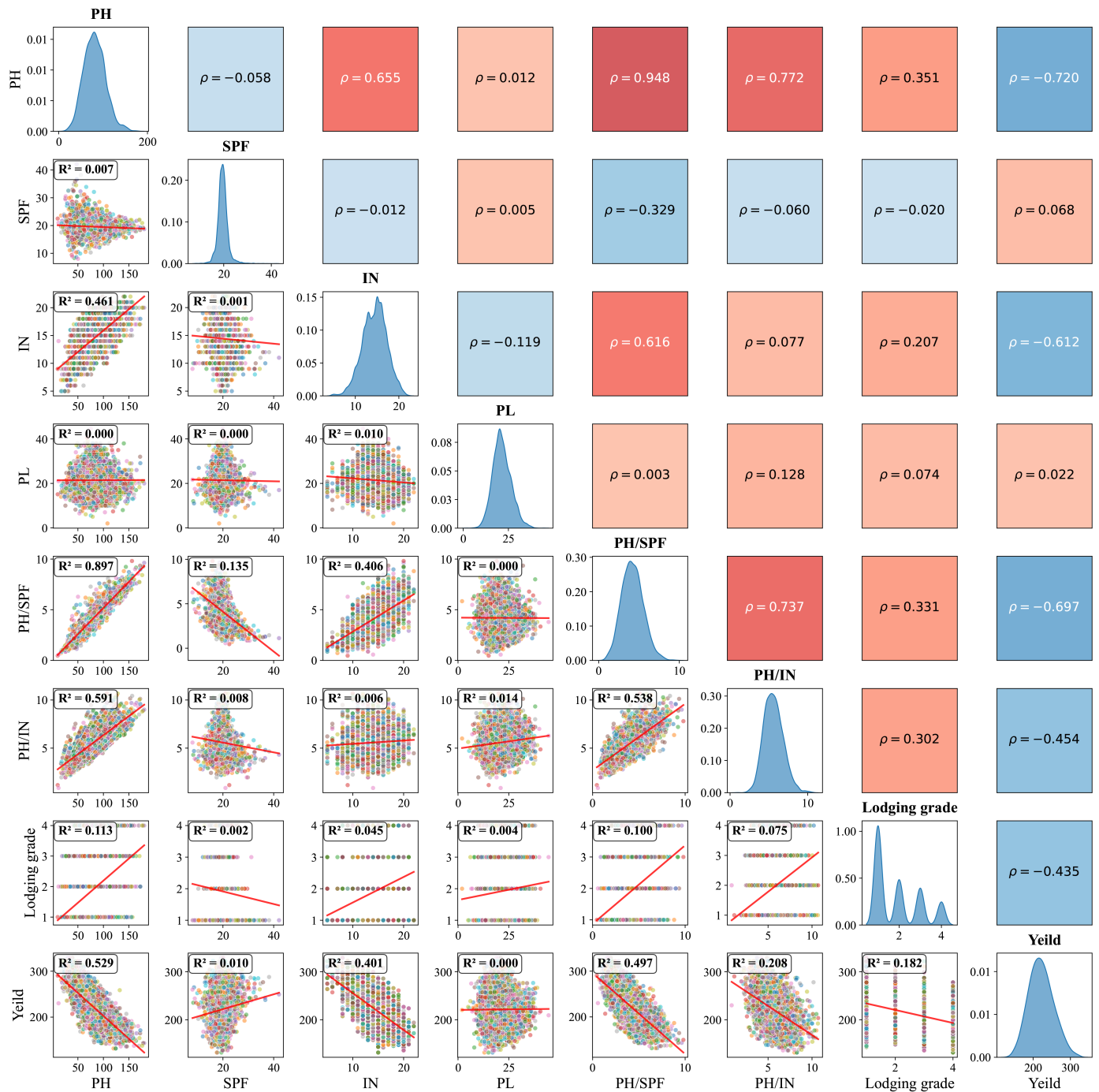


Figure 5. Pairwise correlation matrix of morphological traits, stem mechanical traits, lodging grade and grain yield. The diagonal panels present frequency distribution histograms for all measured soybean traits from 2024 to 2025. The lower triangle contains pairwise scatter plots overlaid with linear fitting curves and the corresponding coefficients of determination R^2 , the red lines represent linear regression fitting lines, and colored dots denote different experimental groups. The upper triangle reports Spearman's rank correlation coefficients ρ , red background indicates positive correlation, and blue background indicates negative correlation; higher color saturation represents stronger correlation magnitude. PH, plant height; SPF, stem pulling force; IN, internode number PL, petiole length of the uppermost third leaf; PH/SPF, plant height-to-stem pulling force ratio; PH/IN, plant height-to-internode number ratio; Lodging grade, lodging grade; Yield, yield per unit area.

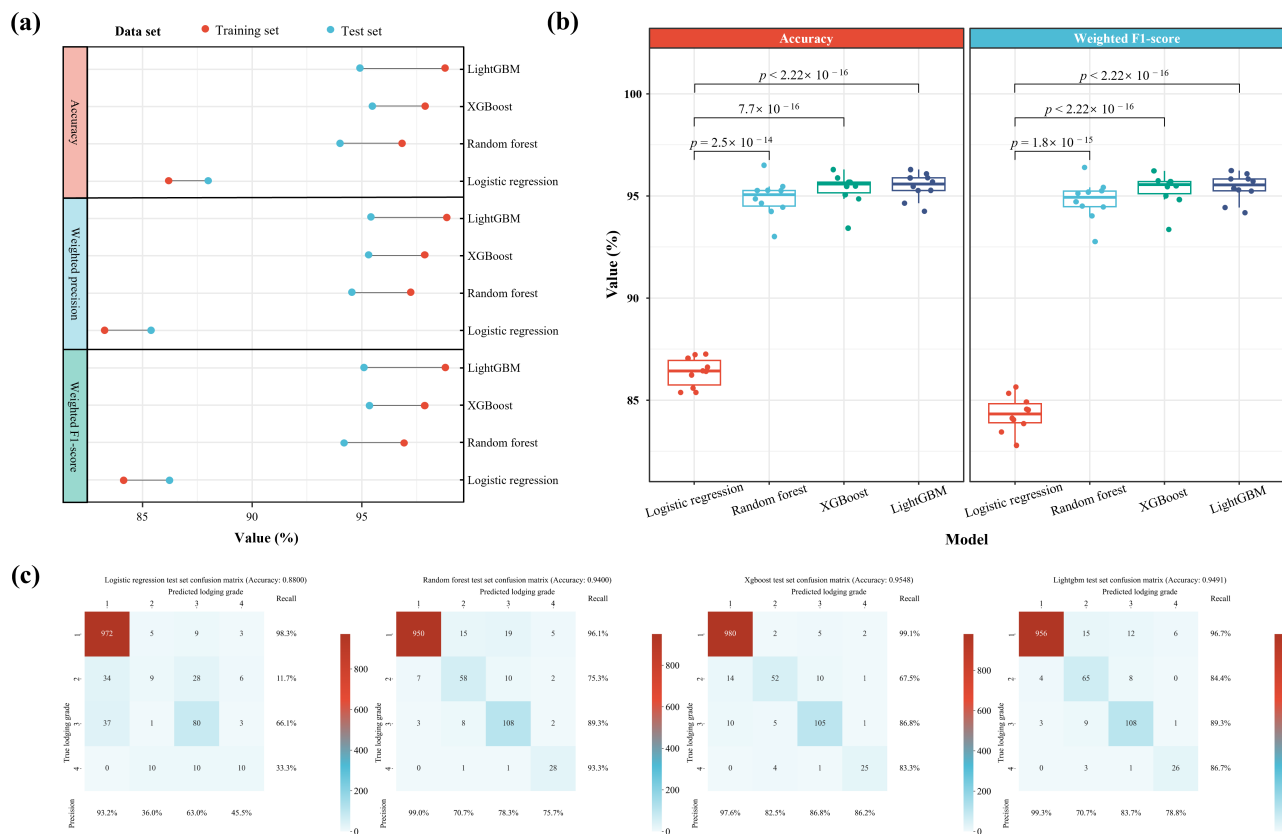


Figure 6. Comprehensive comparison of prediction performance across four models. (a) Comparison of accuracy, weighted precision and weighted F1-score between the training and test sets. Red dots denote the training set, blue dots denote the test set, and the length of the connecting line between paired dots reflects the performance gap between the training and test sets. (b) Boxplots of accuracy and weighted F1-score obtained from stratified 10-fold cross-validation, together with pairwise significance tests among models. (c) Confusion matrices of each model on the test set, with per-class precision and recall values annotated simultaneously, and the color gradient from blue (low) to red (high) indicates the number of classified samples.

Figure 6a depicts the distribution of each evaluation metric across the training and test sets. Overall, the ensemble tree-based models consistently outperformed logistic regression in prediction performance. For logistic regression, the test set accuracy, weighted precision and weighted F1-score were 0.8800, 0.8539 and 0.8623, respectively. As a linear classifier, it was insufficient to fully characterize the nonlinear relationships between morphological and stem mechanical traits and lodging grade, resulting in constrained fitting capability. Random forest substantially improved prediction accuracy, with a test set accuracy of 0.9400 and a weighted F1-score of 0.9419, representing a 9.2% improvement over logistic regression. This finding underscores the strength of ensemble learning in addressing nonlinear classification tasks. The two gradient-boosted tree models achieved the strongest overall performance: XGBoost and LightGBM attained test set accuracies of 0.9548 and 0.9491, with weighted F1-scores of 0.9534 and 0.9509, respectively. Among them, XGBoost delivered the optimal comprehensive performance on the test set, as its built-in regularization mechanism renders it more compatible with the imbalanced multi-class classification setting of this study. Comparison of metric discrepancies between the training and test sets revealed no pronounced overfitting for logistic regression. All three ensemble tree-based models exhibited mild overfitting, whereas XGBoost had the smallest accuracy difference between the training and test sets (-0.024), demonstrating superior generalization stability relative to random forest and LightGBM.

Model prediction stability was evaluated using stratified 10-fold cross-validation in conjunction with paired *t*-tests (Figure 6b). Consistent trends were observed for both accuracy and weighted F1-score: the three ensemble tree-based models significantly outperformed logistic regression. No statistically significant differences in overall performance were detected among the three tree-based models; nevertheless, XGBoost exhibited a more compact boxplot distribution with a higher median value and the lowest dispersion across cross-validation folds, indicating superior generalization robustness. The weighted F1-score balances precision and recall across all classes, making it a more appropriate metric for an imbalanced lodging grade datasets. Post hoc tests revealed a significant difference between random forest and XGBoost in terms of balanced multi-class prediction capacity.

The confusion matrix provides a direct visualization of misclassification patterns for each lodging grade (Figure 6c). Logistic regression achieved acceptable identification performance exclusively for Grade 1 (mild lodging) samples, but yielded low precision and recall for Grade 2 and Grade 4 lodging, and failed to effectively discriminate between intermediate lodging grades. All three ensemble tree-based models delivered uniformly improved recognition accuracy across all grades. XGBoost maintained consistently high and balanced precision and recall across all four lodging grades. While random forest and LightGBM performed well in identifying extreme lodging grades (mild and severe), their classification accuracy for intermediate grades was inferior to that of XGBoost.

Collectively, with respect to prediction accuracy, generalization stability and balanced multi-class classification performance, XGBoost achieved the best overall performance under the current parameter configurations and was thus selected as the optimal model for soybean lodging grade prediction in this study.

3.5. Mechanistic Dissection of Key Lodging-Regulating Traits in Soybean Based on the Optimal Prediction Model

To decipher the interpretability of the optimal XGBoost lodging grade prediction model, the SHAP framework was applied to quantify the contribution magnitude, effect direction and nonlinear response patterns of morphological and stem mechanical traits (Figure 7). Three global sensitivity approaches—perturbation analysis, Sobol variance decomposition and Morris screening—were integrated for cross-validation to systematically elucidate the regulatory roles of each trait in lodging grading.

Globally (Figure 7a), the plant height-to-node number ratio (PH/IN), plant height and plant height-to-stem pulling force ratio (PH/SPF) were identified as the three core driving indicators, all positively correlated with lodging risk. Internode number and petiole length made moderate contributions, while stem pulling force was the only trait exerting a negative protective effect. The two composite indicators exhibited substantially stronger global explanatory power for lodging grading than any single morphological or mechanical trait.

All three core indicators displayed distinct critical risk thresholds (Figure 7b). Below the thresholds, their SHAP values remained negative and suppressed lodging risk; once the thresholds were exceeded, SHAP values rapidly turned positive and sharply increased lodging risk. Synergistic interactions were observed: high plant height combined with low stem pulling force or abundant nodes markedly amplified lodging risk, confirming that no single trait independently determines lodging grade.

Stratified analysis revealed a clear gradient shift in dominant regulatory traits across lodging severity levels (Figure 7c). Mild lodging was dominated by an imbalanced height-to-node ratio, with stem pulling force still retaining notable protective effects. As lodging severity escalated, the contributions of absolute morphological traits (plant height and node number) increased continuously, while the protective effect of stem pulling force diminished progressively. For severe lodging, plant height and node number became the

top two dominant factors, as extreme morphological load overwhelmed the mechanical support capacity of stems.

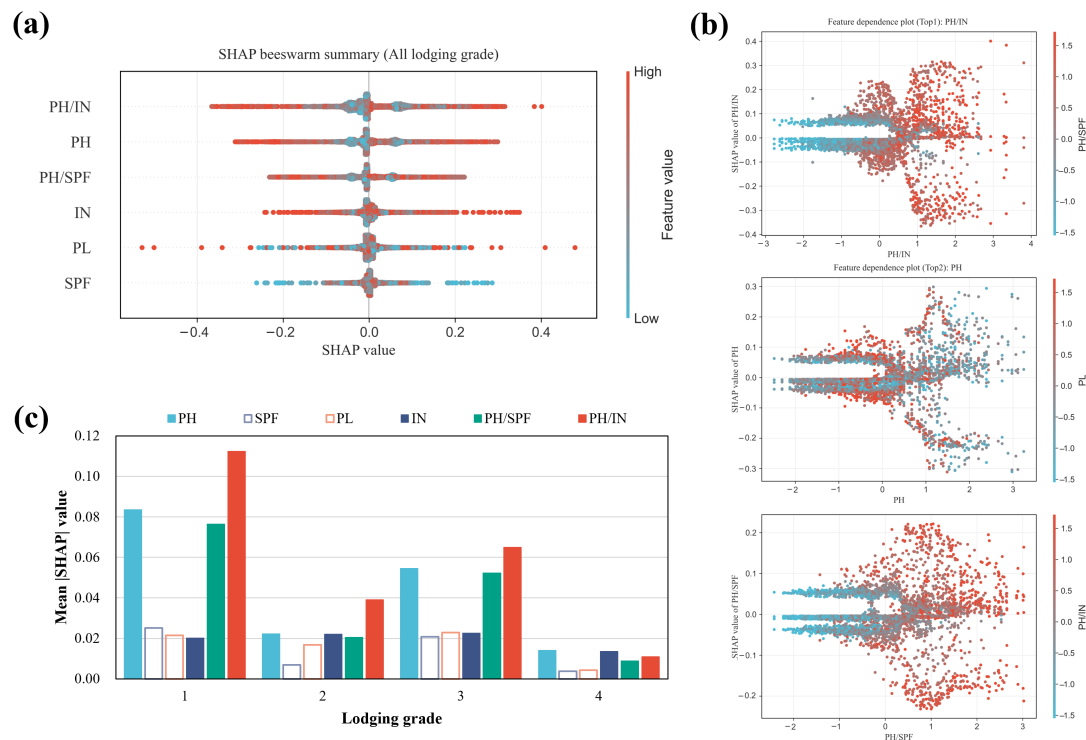


Figure 7. SHAP-based feature interpretation of the XGBoost model. (a) Global feature effect and importance analysis. (b) Univariate effect analysis of core features. (c) Gradient heterogeneity of feature contributions across different lodging grades. Point color indicates the relative magnitude of the feature value from low (blue) to high (red). PH, plant height; SPF, stem pulling force; IN, internode number; PL, petiole length of the uppermost third leaf; PH/SPF, plant height-to-stem pulling force ratio; PH/IN, plant height-to-internode number ratio.

To quantify the variance contribution of each trait to model outputs and validate SHAP findings, three global sensitivity approaches—perturbation analysis, Sobol variance decomposition and Morris screening—were implemented for cross-validation (Figure 8).

Perturbation analysis revealed that single morphological traits (petiole length and internode number) caused the most prominent prediction fluctuations (Figure 8a). Sobol analysis showed (Figure 8b) consistently higher total effect indices (ST) than first-order indices (S1) across all traits, confirming strong inter-trait interactions; PH/IN, IN and PH exhibited the highest total effects, with composite ratio traits deeply involved in multi-factor interactions. Morris screening (Figure 8c) identified PH/IN and PH/SPF as core sensitive traits with both high main effects (μ^*) and interaction intensities (σ), while stem pulling force showed the weakest main effect. Overall, results aligned across methods (Figure 8d): PH/IN served as a stable core driver, and composite ratio traits outperformed single traits in global explanatory power, with stem pulling force contributing the least.

Integrative analysis identified PH/IN, PH and PH/SPF as the three core lodging-driving indicators, all positively correlated with lodging risk, whereas stem pulling force is the sole protective trait. Composite ratio traits far outperformed single traits in explanatory power. All core traits exhibited distinct risk thresholds, with strong synergistic interactions amplifying lodging risk. Dominant regulators shift from height-node ratio imbalance in mild lodging to excessive absolute morphological traits in severe lodging, with a progressive attenuation in stem pulling force's protective effect. Cross-validated by multiple sensitivity methods, PH/IN is a stable core driver.

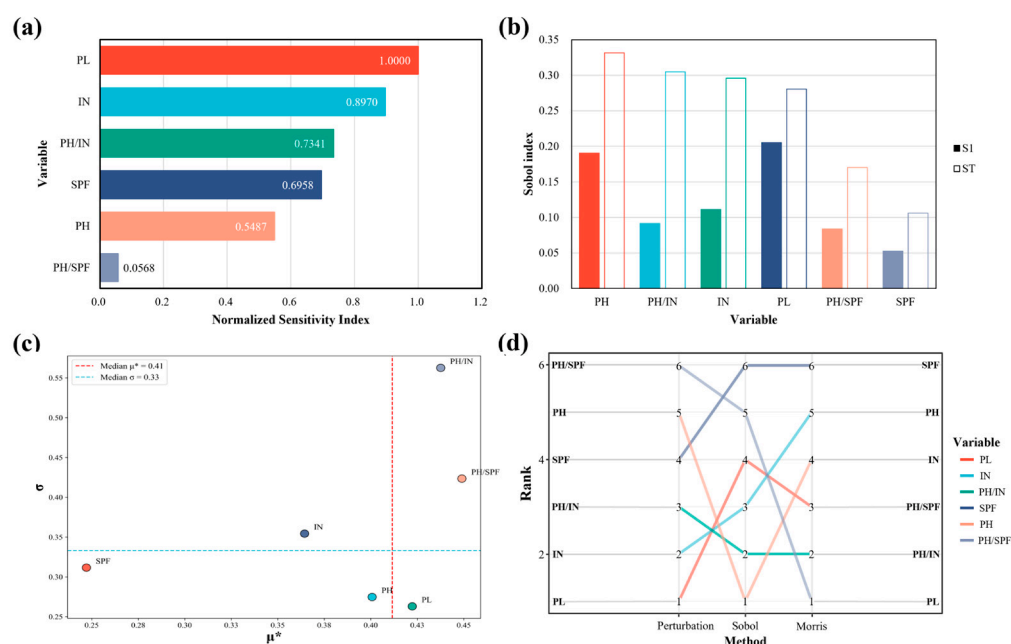


Figure 8. Global sensitivity analysis of the XGBoost model. (a) Perturbation sensitivity analysis; (b) Sobol global sensitivity analysis; (c) Morris screening method; (d) Comprehensive comparison of three global sensitivity approaches. PL, petiole length; IN, internode number; PH/IN, plant height-to-internode number ratio; SPF, stem pulling force; PH, plant height; PH/SPF, plant height-to-stem pulling force ratio.

4. Discussion

4.1. Comparative Evaluation of Prediction Models and the Lodging–Yield Relationship

Accurate lodging grade prediction is central to mitigating yield losses in intensive soybean production. Compelling evidence indicates that lodging events between the pod-setting and seed-filling stages can substantially reduce soybean yield [3,4]. Consistently, our two consecutive years of field data revealed a significant negative correlation between lodging grade and grain yield, with yield declining continuously as lodging severity increased. Lodging grade explained a substantial proportion of total yield phenotypic variation and was identified as a dominant driver of yield differentiation among tested germplasm. For the high-density mulched drip irrigation system, where lodging is a persistent bottleneck to stable yield growth, a high-precision multi-class lodging prediction model provides essential quantitative support for lodging-resistant germplasm screening and targeted agronomic regulation.

This study compared the predictive efficacy of one linear classifier and three ensemble learning frameworks for ordinal lodging classification across multiple severity levels. Logistic regression achieved the lowest test-set accuracy among the four candidate models. This inferior performance is fundamentally attributed to the inherent monotonic mapping assumption embedded in linear classifiers [27], which cannot resolve trait multicollinearity or capture nonlinear threshold effects from trait interactions, imposing an inherent accuracy bottleneck for multi-level ordinal classification. In contrast, gradient-boosted tree models delivered markedly superior classification performance. This corroborates that soybean lodging, as a complex trait shaped by multiple interacting factors [28], cannot be accurately delineated by simple linear relationships [9], and ensemble learning approaches are better suited to predicting such complex agronomic traits [29].

XGBoost consistently outperformed the other evaluated models in classification accuracy, overall fitting quality and overfitting control. Its superior performance arises from three core algorithmic features aligned with the study objectives [30]. First, second-order

Taylor expansion of the loss function improves quantification of classification boundary gains and reduces cross-grade misclassification relative to first-order gradient boosting [31]. Second, the built-in multi-dimensional regularization system effectively constrains model complexity and suppresses overfitting under the moderate sample sizes typical of agricultural field trials [32], as evidenced by its smallest performance gap between the training and test sets. Third, dual random sampling across observations and features enhances model robustness against sample noise derived from field measurement errors and environmental fluctuations [33]. The bagging-based random forest exhibited weaker capacity to fit nonlinear trait coupling relationships than boosting-based models [34]. These conclusions are statistically validated by stratified 10-fold cross-validation and paired *t*-tests.

The performance pattern of models in this study is generally consistent with previous machine learning-based lodging research in other crops, while also showing unique application advantages. For soybean lodging assessment, Sarkar et al. [4] achieved high classification accuracy using UAV remote sensing imagery and artificial neural networks, but their method relies on specialized imaging equipment and is susceptible to weather conditions, limiting large-scale and low-cost application in breeding nurseries. In contrast, our model adopts four easily measurable field traits as inputs, which greatly reduces the application threshold and improves operability for early-generation germplasm screening. In maize lodging research, Huang et al. [9] reported that ensemble learning algorithms significantly outperformed linear models in capturing nonlinear interactions among lodging-related traits, which is highly consistent with our finding that XGBoost achieved the optimal prediction accuracy and generalization stability. In addition, compared with studies using single indicators such as plant height or stem strength for lodging evaluation [7,35], the composite indices PH/IN and PH/SPF constructed in this study exhibited stronger discriminatory power, which supports the conclusion that comprehensive indicators integrating load effect and support effect can better reflect the actual lodging resistance of crops.

4.2. Key Determinant Traits and Grading Heterogeneity Patterns of Lodging Grade

Building on the identification of XGBoost as the optimal lodging grade prediction model, this study further dissected the key regulatory traits and grading heterogeneity of soybean lodging under mulched drip irrigation via cross-validation using SHAP interpretability analysis combined with three global sensitivity methods: perturbation analysis, Sobol variance decomposition and Morris screening. The plant height-to-internode number ratio, plant height, and plant height-to-stem pulling force ratio were identified as three core predictive traits. Composite ratio traits showed substantially stronger explanatory power for lodging classification than single traits, confirming that soybean lodging is an integrated phenotype modulated by multi-dimensional trait interactions rather than a simple trait determined by a single factor [28].

Mechanistically, the three core factors regulate plant stability through distinct functional dimensions. Plant height serves as the fundamental morphological load factor, directly determining the vertical position of the plant's center of gravity [36]. The bending moment borne by the basal stem is positively correlated with the gravity center height: taller plants sustain greater bending loads under external forces and correspond to a lower critical lodging threshold [37]. Under super-high-yield mulched drip irrigation system, high planting density and adequate water and fertilizer readily induce internode elongation and increased plant height [38], while canopy closure inhibits stem mechanical tissue development and further amplifies lodging risk [39]. In this study, plant height exhibited the highest global contribution among all basic traits, consistent with conclusions from previous soybean lodging research [40].

Plant height alone cannot reflect intrinsic stem mechanical strength, and isolated stem pulling force overlooks bending moment variations caused by plant morphology [41]. The plant height-to-stem pulling force ratio integrates the load effect of plant height and the support effect of stem mechanical strength [42], effectively circumventing the one-sidedness of single-trait evaluation and enabling more accurate characterization of comprehensive lodging resistance across diverse genotypes and cultivation conditions. From a structural mechanics view, more main stem internodes increase the effective load-bearing length of stems, significantly reducing flexural capacity and critical buckling resistance [35]. Increased internode count usually accompanies elevated plant height and self-weight, which superimpose with external loads to aggravate stem mechanical burden [43]. The plant height-to-internode number ratio ranks among the top traits in predictive contribution, particularly dominating mild lodging identification, while single internode count shows notably weaker explanatory power despite a stable increasing trend with lodging grade.

Grade-specific analysis revealed a clear gradient shift in dominant regulatory traits as lodging severity escalates. In mild lodging, the plant height-to-internode number ratio acts as the leading determinant, with stem pulling force still exerting a notable protective effect. As lodging worsens, the weights of absolute morphological traits rise continuously while stem pulling force's protective effect diminishes. In severe lodging, plant height and internode count become core drivers, as stem mechanical support can no longer offset the load from extreme plant architecture, explaining the sharp escalation of lodging risk in super-high-yield cultivation [44,45]. Furthermore, traits interact synergistically rather than acting independently. Sobol analysis showed total sensitivity indices far exceeding first-order indices for all traits, and such nonlinear interactions underpin the limited accuracy of single-trait evaluation frameworks.

4.3. Generalization Stability of the Model

Generalization stability is a fundamental prerequisite for translating predictive models from experimental research to field-scale applications [46]. In terms of overall stability performance, XGBoost exhibited substantially superior generalization robustness relative to the other three comparative models. The 10-fold stratified cross-validation results revealed that XGBoost presented the narrowest distribution range of accuracy values and the highest median accuracy, with the lowest degree of performance variability across folds among the four models. This finding confirms that the model has stronger resilience to sampling fluctuations and does not suffer from dramatic oscillations in prediction accuracy caused by compositional differences in training subsets. With respect to overfitting control, the accuracy gap between the training and test sets for XGBoost was merely 0.024, the smallest value across all reference algorithms. Its built-in multi-dimensional regulatory mechanisms, including maximum tree depth constraints, weight regularization and minimum split gain thresholds, effectively constrain model complexity, prevent overfitting to idiosyncratic noise in the training set, and algorithmically underpin the output stability of the model when applied to unseen samples [47]. Furthermore, all core input traits of the model, namely plant height, main stem internode number, petiole length of the third uppermost leaf and stem pulling force, are quantitative morphological and mechanical indicators that can be directly measured under field conditions. The model exhibits high tolerance to minor measurement errors introduced by routine field operations, and avoids abrupt misclassification of lodging grades due to measurement deviations, which further enhances the reliability of prediction results in practical field deployment.

4.4. Limitations and Future Research Perspectives

Despite the favorable prediction performance and clear regulatory mechanism revealed in this study, several limitations still exist that should be acknowledged. First, all field trials were conducted at a single experimental station (Shihezi, 85.9800 °E, 44.3160 °N) under mulched drip-irrigated high-density soybean cultivation systems; thus, the generalizability of the proposed model to other ecological regions, soil types and planting patterns remains to be further verified. Second, although two consecutive years of phenotypic data were collected to account for inter-annual environmental variation, the dataset still cannot fully cover extreme climatic scenarios such as persistent rainstorms and strong winds, which may affect the robustness of the model under extreme lodging conditions.

For future research, multi-site joint trials across different ecological regions are recommended to further calibrate the model and improve its cross-regional generalization ability. Meanwhile, integrating multi-omics data and meteorological factors into the prediction framework will help to build a more comprehensive lodging risk early warning system for field production.

5. Conclusions

To address the critical practical bottlenecks of frequent soybean lodging, insufficient accuracy of conventional evaluation systems, and the lack of feasible non-destructive pre-identification approaches under high-density super-high-yield cultivation with mulched drip irrigation, this study developed and validated a non-destructive prediction model for soybean lodging grades based on multi-dimensional phenotypic traits, and systematically elucidated the regulatory patterns of key traits underlying lodging risk. The results demonstrate that soybean lodging performance is jointly governed by morphological traits, mechanical traits and their derived composite indicators. All core traits exhibit significant associations with lodging grade, among which plant height, the plant height-to-stem pulling force ratio and the plant height-to-internode number ratio constitute the core predictive feature set for lodging grade classification. Composite ratio traits substantially outperform individual basic traits in overall explanatory power for lodging risk. By effectively mitigating the inherent limitations of univariate assessment, they serve as robust comprehensive indicators for population-level lodging risk evaluation in soybean. Among the four constructed prediction models, the XGBoost model achieves the best overall performance, and effectively addresses the long-standing limitations of low classification accuracy and strong subjectivity in traditional soybean lodging resistance research. The non-destructive lodging risk prediction framework established in this study provides a quantitative foundation and technical support for accurate and convenient field lodging risk identification, cultivation management optimization, and lodging-resistant high-yield soybean breeding.

Author Contributions: Conceptualization, X.Z.; methodology, X.Z., J.Z., Y.G., Z.Z. and K.Z.; investigation, X.Z., X.S., H.Z., J.Z., Y.G., Z.Z. and K.Z.; data curation, X.S. and H.Z.; formal analysis, X.Z., X.S. and H.Z.; software, X.Z.; validation, X.S., H.Z., J.Z., Y.G., Z.Z., K.Z. and Y.Z.; writing—original draft preparation, X.Z.; writing—review and editing, J.Z. and X.Z.; supervision, Y.Z.; project administration, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Bingtuan Science and Technology Program (2024YJQN01), National Key Research and Development Program of China: Major Project on Agricultural Bio-breeding (2024ZD040780305), Youth Science Foundation Project of the Xinjiang Production and Construction Corps (2024DB014), and Shihezi Comprehensive Experimental Station, National Soybean Industrial Technology System.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: In this study, a generative artificial intelligence large language model was employed as an auxiliary tool for manuscript polishing and machine learning model code implementation. In the writing stage, it was used exclusively for refining academic expressions and optimizing textual presentation to enhance manuscript readability and linguistic consistency; in the coding stage, it supported classification model framework construction, syntax correction and code formatting standardization. Rigorous process controls were implemented to uphold research originality and academic integrity. All AI-polished content has been fully reviewed, manually revised and verified paragraph by paragraph by the authors to ensure alignment with field measurement data, experimental design and core findings. All AI-assisted code segments have been manually debugged and validated against the measured dataset, with no alteration to the study's score algorithmic logic and analytical conclusions. The authors retain full authority over the final manuscript, model codes and research conclusions, assume full responsibility for all content of this paper, and this disclosure complies with relevant requirements for academic transparency.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

PH	Plant height
SPF	Stem pulling force
IN	Internode number
PL	Petiole length
PH/IN	Plant height-to-internode number ratio
PH/SPF	Plant height-to-stem pulling force ratio

Appendix A

Table A1. List of soybean cultivars (lines) examined in this study.

Number	Material Name	Number	Material Name	Number	Material Name	Number	Material Name
1	12K2033	2	12K2035	3	12K2040	4	12K2055
5	12K2066	6	12K2090	7	12K2092	8	12K2108
9	12K2113	10	12K2162	11	12K2163	12	12K2173
13	12K2173-1	14	12K2198	15	12K2231-1	16	16NF5239
17	17-157	18	17-164	19	17-189	20	17-294
21	17-323	22	17-448	23	18-21	24	2001-336-7
25	2011-X-479	26	2011-X-639	27	2011-X-647	28	2016MS1
29	2016MS1-1	30	2016MS1-2	31	2016MS1-200	32	2016MS1-3
33	2016MS1-300	34	2016MS1-7	35	2017MS1	36	2017MS1-487
37	2017MS1-488	38	2017MS1-489	39	2017MS1-490	40	2017MS1-491
41	2017MS1F5	42	2017Ms1F5-700	43	2017MS6-321	44	2017MS6-322
45	2017MS6-323	46	2017MS6-324	47	2017MS6-325	48	2017MS6-326
49	2017MS6-327	50	2017MS6F5	51	2017MS6-Y	52	2017MSF3-128-1
53	2017MSF3-128-2	54	2017MSF3-176-1	55	2017MSF3-176-2	56	2017MSF3-190-4
57	2017MSF3-226-1	58	2017MSF3-244-1	59	2017MSF3-244-2	60	2017MSF3-244-3
61	2017MSF3-3-283-1	62	2017MSF3-3-283-2	63	2017MSF3-63-1	64	2017MSF3-63-2
65	312F9-4	66	4Q-1-1	67	4Q-2-1	68	753-1
69	887704-4	70	C1212F8-2-54	71	C13001F8-3-68	72	C1300578-3-16-6
73	C1603F4-15-2	74	C1603F4-5-9-3	75	DA17-5	76	DN50
77	DW080	78	DW083	79	DW143	80	DW267
81	DW270	82	HN106	83	HN108	84	HN113
85	HN126	86	HN141	87	HN144-2	88	HN154
89	HN156	90	HN165	91	HN167	92	HN173
93	HN178	94	HN184	95	HN190	96	HN196
97	HN199	98	HN200	99	HN205	100	HN219
101	HN244	102	HN245	103	HN247	104	HN249
105	HN25	106	HN251	107	HN26	108	HN28
109	HN287	110	HN31	111	HN316	112	HN32

Table A1. Cont.

Number	Material Name	Number	Material Name	Number	Material Name	Number	Material Name
113	HN341-2	114	HN341-4	115	HN348	116	HN360
117	HN40	118	HN408	119	HN443	120	HN47
121	HN477	122	HN481	123	HN485	124	HN491
125	HN492	126	HN494	127	HN50	128	HN507
129	HN56	130	HN59	131	HN66	132	HN67
133	HN84	134	HN97	135	NK0906-19	136	SE070
137	SE118	138	TZX-1171	139	TZX-1171-5-J	140	TZX-1349
141	TZX-1380-1	142	TZX-1426	143	TZX-1498	144	TZX-1559-1
145	TZX-1577-1	146	TZX-1577-2	147	TZX-1599	148	TZX-1815
149	TZX-1826	150	TZX-1834	151	TZX-1846-2	152	TZX-288
153	TZX-5381-1	154	BA2	155	BJ1	156	94-2-3-4-1
157	CD201	158	DCD401	159	DFD1	160	DN60
161	DND251	162	DND305	163	DS37	164	DS112
165	DS29	166	0336-10	167	90-33-17	168	HJ1
169	HJ2	170	HT2	171	HF55	172	HF08-1
173	HN119	174	HN124	175	HND126	176	HN132
177	HN133	178	HN134	179	HN138	180	HN71
181	HN76	182	HN85	183	1428	184	HH45
185	HH50	186	HH53	187	HK57	188	HK60
189	HK79	190	HN508	191	HN54	192	HN69
193	HND71	194	HN80	195	HN81	196	HND84
197	HN86	198	HN87	199	HN88	200	HN90
201	HN93	202	HN98	203	HN417	204	X2001-336-7
205	JD19	206	JD29	207	JL31	208	JN139
209	JN18	210	JN233	211	JN28	212	JY101
213	JY209	214	JY210	215	JY251	216	JY252
217	JY256	218	JY259	219	JY260	220	JY306
221	JY308	222	JY309	223	JY310	224	JY354
225	JY406	226	JY407	227	JY414	228	JY419
229	JY419500	230	JY441	231	JY466	232	JY481
233	JY491	234	JY496	235	JY496500	236	JY498400
237	JY554	238	JY584	239	JY593	240	JY627
241	JY86	242	JY3	243	J25	244	J34
245	J55	246	JY8	247	KX8	248	KD43
249	KD66	250	KD94	251	KF20	252	KF13
253	KN30	254	KN33	255	KYD1	256	LF23
257	LK324	258	LK3425	259	LPD	260	MD1137
261	MD359	262	MD36	263	MD45	264	MD46
265	MD48	266	MD49	267	MD50	268	MD23
269	NHD2	270	NQD20	271	NQD24	272	PD2
273	PD3	274	QN11	275	QN12	276	QN7
277	QXD1	278	12-550	279	15-712	280	17-171
281	18-3106	282	20-1241	283	20-1509	284	XD8-3
285	SD15	286	SD45	287	SDD2-15	288	SDD2-40
289	SHD1	290	SN14	291	SN35	292	SN36
293	SN42	294	SN52	295	SN53	296	SN62
297	SN71	298	SN76	299	SN81	300	SN82
301	SN94	302	SW2	303	SW3	304	TD63
305	TN61	306	20-6-7	307	XD11	308	XD13
309	XD14	310	XD1	311	XD22	312	XD23-38
313	XD23	314	XD25	315	XD26	316	XD28
317	XD2	318	XD3	319	XD8	320	XD9
321	XND2	322	XND3	323	XND5	324	XND9
325	XZD10	326	XZD11	327	XZD12	328	XZD13
329	XZD14	330	XZD15	331	XZD1	332	XZD6
333	XZD7	334	XZD8	335	XZD9	336	2011-MEB9
337	CD30	338	CD6	339	CN301	340	CN306
341	CN35	342	CN39	343	CN63	344	CN64
345	ZH35	346	ZH45	347	ZH901	348	ZH902
349	ZH909	350	ZH911	351	ZJ602	352	ZK42
353	ZL102	354	ZL606	355	ZL608	356	ZL1

References

1. Wilson, R.F. Outlook for soybeans and soybean products in 21st century markets. *Lipid Technol.* **2010**, *22*, 199–202. [[CrossRef](#)]
2. Li, S.; Sun, Z.; Sang, Q.; Qin, C.; Kong, L.; Huang, X.; Liu, H.; Su, T.; Li, H.; He, M.; et al. Soybean reduced internode 1 determines internode length and improves grain yield at dense planting. *Nat. Commun.* **2023**, *14*, 7939. [[CrossRef](#)] [[PubMed](#)]

3. Ye, Y.; Cheng, Z.; Yang, X.; Yang, S.; Tang, K.; Yu, H.; Gao, J.; Zhang, Y.; Leng, J.; Zhang, W.; et al. Lrm3 positively regulates stem lodging resistance by degrading myb6 transcriptional repressor in soybean. *Plant Biotechnol. J.* **2025**, *23*, 2978–2993. [[CrossRef](#)] [[PubMed](#)]
4. Sarkar, S.; Zhou, J.; Scaboo, A.; Zhou, J.; Aloysius, N.; Lim, T.T. Assessment of soybean lodging using uav imagery and machine learning. *Plants* **2023**, *12*, 2893. [[CrossRef](#)] [[PubMed](#)]
5. Gao, H.-W.; Yang, M.-Y.; Yan, L.; Hu, X.-Z.; Hong, H.-L.; Zhang, X.; Sun, R.-J.; Wang, H.-R.; Wang, X.-B.; Liu, L.-K.; et al. Identification of tolerance to high density and lodging in short petiolate germplasm M657 and the effect of density on yield-related phenotypes of soybean. *J. Integr. Agric.* **2022**, *22*, 434–446. [[CrossRef](#)]
6. Li, Z.; Liu, F.; Wu, W. Optimising nitrogen management strategies to minimise lodging risk while sustaining high seed yield in rapeseed. *Eur. J. Agron.* **2022**, *142*, 126671. [[CrossRef](#)]
7. Wang, Q.; Xue, J.; Chen, J.-L.; Fan, Y.-H.; Zhang, G.-Q.; Xie, R.-Z.; Ming, B.; Hou, P.; Wang, K.-R.; Li, S.-K. Key indicators affecting maize stalk lodging resistance of different growth periods under different sowing dates. *J. Integr. Agric.* **2020**, *19*, 2419–2428. [[CrossRef](#)]
8. Gebremeskel, K.; Birhane, E.; Tadele, Z.; Habtu, S.; Haile, M.; Chanyalew, S.; Assefa, K. Arbuscular mycorrhizal fungi reduce lodging of teff [*Eragrostis tef* (Zucc.) Trotter] under N fertilizer applications. *Arid. Land Res. Manag.* **2025**, *39*, 578–601. [[CrossRef](#)]
9. Huang, G.; Zhang, Y.; Gu, S.; Wen, W.; Lu, X.; Guo, X. Identifying key factors influencing maize stalk lodging resistance through wind tunnel simulations with machine learning algorithms. *Artif. Intell. Agric.* **2025**, *15*, 316–326. [[CrossRef](#)]
10. Yang, J.; Li, M.; Yin, Y.; Liu, Y.; Gan, X.; Mu, X.; Li, H.; Li, J.; Li, H.; Zheng, J.; et al. Spatial accumulation of lignin monomers and cellulose underlying stalk strength in maize. *Plant Physiol. Biochem.* **2024**, *214*, 108918. [[CrossRef](#)] [[PubMed](#)]
11. Liu, W.; Deng, Y.; Hussain, S.; Zou, J.; Yuan, J.; Luo, L.; Yang, C.; Yuan, X.; Yang, W. Relationship between cellulose accumulation and lodging resistance in the stem of relay intercropped soybean [*Glycine max* (L.) merr.]. *Field Crop. Res.* **2016**, *196*, 261–267. [[CrossRef](#)]
12. Yu, J.; Zhang, R.; Ma, X.; Jia, Z.; Li, X.; Li, Y.; Liu, H. Physiological mechanism of lodging resistance of oat stalk and analysis of transcriptome differences. *Front. Plant Sci.* **2025**, *16*, 1532216. [[CrossRef](#)] [[PubMed](#)]
13. Rehman, M.; Luo, D.; Mubeen, S.; Pan, J.; Cao, S.; Saeed, W.; Chen, P. Progress in agronomic crops lodging resistance and prevention: A review. *J. Agron. Crop Sci.* **2024**, *210*, e12785. [[CrossRef](#)]
14. Huang, G.; Li, Y.; Zhang, Y.; Wen, W.; Zhao, C.; Guo, X. Overcoming challenges in plant biomechanics: Methodological innovations and technological integration. *Adv. Sci.* **2025**, *12*, e2415606. [[CrossRef](#)] [[PubMed](#)]
15. Luo, H.; Luo, R.; Liu, T.; Guo, W.; Zeng, Z.; Zhu, B. Progress, challenges, and strategies in lodging resistance research of gramineous crops. *Sci. Agric. Sin.* **2026**, *59*, 1587–1607.
16. Zheng, Y.; Hou, P.; Jia, X.; Zhu, L.; Zhao, Y.; Song, W.; Song, W.; Guo, J. Evaluation of the lodging resistance and the selection of identification indexes of maize inbred lines. *Food Energy Secur.* **2023**, *12*, e499. [[CrossRef](#)]
17. Pyne, S.; Negi, M.; Mahapatra, B.; Kumar, A.; Dey, P.; Pandit, P. Hybrid modeling and management of lodging in subtropical fiber flax: Optimizing plant nutrition for sustainable cultivation. *J. Nat. Fibers* **2024**, *21*, 2394139. [[CrossRef](#)]
18. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [[CrossRef](#)] [[PubMed](#)]
19. Redhu, N.S.; Thakur, Z.; Yashveer, S.; Mor, P. Artificial intelligence: A way forward for agricultural sciences. In *Bioinformatics in Agriculture: Next Generation Sequencing Era*; Academic Press: Cambridge, MA, USA, 2022; pp. 641–668. [[CrossRef](#)]
20. Mourtzinis, S.; Esker, P.D.; Specht, J.E.; Conley, S.P. Advancing agricultural research using machine learning algorithms. *Sci. Rep.* **2021**, *11*, 17879. [[CrossRef](#)] [[PubMed](#)]
21. Wang, L.; Cheng, B.; Zhou, T.; Jing, S.; Liu, R.; Gao, Y.; Deng, C.; Ye, W.; Luo, Z.; Raza, A.; et al. Quantifying the effects of plant density on soybean lodging resistance and growth dynamics in maize-soybean strip intercropping. *Front. Plant Sci.* **2023**, *14*, 1264378. [[CrossRef](#)] [[PubMed](#)]
22. Qu, X.; Zhou, J.; Gu, X.; Wang, Y.; Sun, Q.; Pan, Y. Monitoring maize lodging severity based on multi-temporal sentinel-1 images using time-weighted dynamic time warping. *Comput. Electron. Agric.* **2023**, *215*, 108365. [[CrossRef](#)]
23. Guo, C.; Tan, F. Swrd-yolo: A lightweight instance segmentation model for estimating rice lodging degree in uav remote sensing images with real-time edge deployment. *Agriculture* **2025**, *15*, 1570. [[CrossRef](#)]
24. Liu, Z.; Li, H.; Fan, X.; Huang, W.; Yang, J.; Zheng, Y.; Wen, Z.; Li, Y.; Wang, D.; Wang, S.; et al. Selection of soybean elite cultivars based on phenotypic and genomic characters related to lodging tolerance. *Plant Breed.* **2017**, *136*, 526–538. [[CrossRef](#)]
25. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
26. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. Sensitivity analysis: From theory to practice. In *Global Sensitivity Analysis: The Primer*; John Wiley & Sons: Hoboken, NJ, USA, 2007; pp. 237–275.
27. Campbell, M.K.; Donner, A. Classification efficiency of multinomial logistic regression relative to ordinal logistic regression. *J. Am. Stat. Assoc.* **1989**, *84*, 587–591. [[CrossRef](#)]

28. Chen, B.; Chai, C.; Duan, M.; Yang, X.; Cai, Z.; Jia, J.; Xia, Q.; Luo, S.; Yin, L.; Li, Y.; et al. Identification of quantitative trait loci for lodging and related agronomic traits in soybean (*Glycine max* [L.] merr.). *BMC Genom.* **2024**, *25*, 900. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Kunduru, B.; Bokros, N.T.; Tabaracci, K.; Kumar, R.; Brar, M.S.; Stubbs, C.J.; Oduntan, Y.; DeKold, J.; Bishop, R.; Woomer, J.; et al. High-density phenotype data of intermediate phenotypes associated with stalk lodging resistance in maize. *Sci. Data* **2025**, *12*, 1721. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Shang, X.; Cui, C.; Sun, X.; Wang, X.; Zhang, J. Classification task-driven hyperspectral band selection via interpretability from xgboost. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 13733–13754. [\[CrossRef\]](#)
31. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T.; et al. *Xgboost: Extreme Gradient Boosting, R Package Version 0.4-2*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
32. Han, L.; Yang, G.; Yang, X.; Song, X.; Xu, B.; Li, Z.; Wu, J.; Yang, H.; Wu, J. An explainable xgboost model improved by smote-enn technique for maize lodging detection based on multi-source unmanned aerial vehicle images. *Comput. Electron. Agric.* **2022**, *194*, 106804. [\[CrossRef\]](#)
33. Anju; Hazarika, A.V. Extreme gradient boosting using squared logistics loss function. *Int. J. Sci. Dev. Res.* **2017**, *2*, 54–61.
34. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [\[CrossRef\]](#)
35. Zhang, Y.; Wang, Y.; Ye, D.; Wang, W.; Qiu, X.; Duan, L.; Li, Z.; Zhang, M. Ethephon improved stalk strength of maize (*Zea mays* L.) mainly through altering internode morphological traits to modulate mechanical properties under field conditions. *Agronomy* **2019**, *9*, 186. [\[CrossRef\]](#)
36. Wang, Y.; Pan, Y.; Zhao, F.; Meng, X.; Li, Q.; Huang, Y.; Ye, Y. Changes in the lodging resistance of winter wheat from 1950s to the 2020s in henan province of china. *BMC Plant Biol.* **2023**, *23*, 442. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Wu, M.; Jiang, H.; Wei, Z.; Li, W.; Gao, K.; Wang, D.; Wei, X.; Tian, P.; Cui, J.; Di, Y.; et al. Influence of nitrogen application rate on stem lodging resistance rice under dry cultivation. *Agronomy* **2023**, *13*, 426. [\[CrossRef\]](#)
38. Zhao, J.; Geng, W.; Xue, Y.; Alam, S.; Liu, P.; Zhao, B.; Ren, B.; Zhang, J. Optimizing plant morphology to enhance canopy light distribution improves lodging resistance and grain yield in densely planted maize. *J. Integr. Agric.* **2025**. [\[CrossRef\]](#)
39. Zhai, J.; Ming, B.; Zhang, G.; Zhang, Y.-m.; Cao, Y.; Xie, R.; Wang, K.; Li, S.; Xue, J.; Gou, L. Nitrogen rate, density, and plant architecture jointly determine lodging risk in drip fertigated spring maize. *Ann. Agric. Sci.* **2025**, *70*, 100399. [\[CrossRef\]](#)
40. Liang, Z.; Zhao, W.; Qi, N.; Cao, Y.; Li, R.; Wang, X.; Shi, G.; Xie, N.; Yao, X.; Xie, F. Genome-wide association study of lodging-related agronomic traits in soybean plants. *Ind. Crops Prod.* **2025**, *236*, 121861. [\[CrossRef\]](#)
41. Xue, J.; Gao, S.; Fan, Y.; Li, L.; Ming, B.; Wang, K.; Xie, R.; Hou, P.; Li, S. Traits of plant morphology, stalk mechanical strength, and biomass accumulation in the selection of lodging-resistant maize cultivars. *Eur. J. Agron.* **2020**, *117*, 126073. [\[CrossRef\]](#)
42. Niu, Y.; Chen, T.; Zhao, C.; Zhou, M. Improving crop lodging resistance by adjusting plant height and stem strength. *Agronomy* **2021**, *11*, 2421. [\[CrossRef\]](#)
43. Kaack, K.; Schwarz, K.-U. Morphological and mechanical properties of miscanthus in relation to harvesting, lodging, and growth conditions. *Ind. Crops Prod.* **2001**, *14*, 145–154. [\[CrossRef\]](#)
44. Stubbs, C.J.; Kunduru, B.; Bokros, N.; Verges, V.; Porter, J.; Cook, D.D.; DeBolt, S.; McMahan, C.; Sekhon, R.S.; Robertson, D.J. Moving toward short stature maize: The effect of plant height on maize stalk lodging resistance. *Field Crops Res.* **2023**, *300*, 109008. [\[CrossRef\]](#)
45. Wang, N.; Wang, F.-X.; Shock, C.C.; Meng, C.-B.; Huang, Z.-J.; Gao, L.; Zhao, J.-Y. Evaluating quinoa stem lodging susceptibility by a mathematical model and the finite element method under different agronomic practices. *Field Crops Res.* **2021**, *271*, 108241. [\[CrossRef\]](#)
46. Nikhil, U.V.; Pandiyan, A.M.; Raja, S.P.; Stamenkovic, Z. Machine learning-based crop yield prediction in south india: Performance analysis of various models. *Computers* **2024**, *13*, 137. [\[CrossRef\]](#)
47. Xia, W.; Liu, B.; Xiang, H. Prediction of liquid accumulation height in gas well tubing using integration of crayfish optimization algorithm and xgboost. *Processes* **2024**, *12*, 1788. [\[CrossRef\]](#)

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.