Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Shapley values as metrics for studying genotype-by-environment interaction

Luan Tiago Dos Santos Carbonari

carbonari.luan@gmail.com

State University of Santa Catarina (UDESC) - Lages Campus

Carlos Zacarias Joaquim Junior

State University of Santa Catarina (UDESC) - Lages Campus

Mauro Bitencourt Souza

State University of Santa Catarina (UDESC) - Lages Campus

Altamir Frederico Guidolin

State University of Santa Catarina (UDESC) - Lages Campus

Sydney Antonio Frehner Kavalco

Agricultural Research and Rural Extension Company of Santa Catarina (EPAGRI)

Jefferson Luís Meirelles Coimbra

State University of Santa Catarina (UDESC) - Lages Campus

Research Article

Keywords: phenotypic stability, adaptability, genotype recommendation, plant breeding

Posted Date: July 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10120365/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Shapley values as metrics for studying genotype-by-environment interaction

Luan Tiago dos Santos Carbonari[1*] · Carlos Zacarias Joaquim Junior[1] · Mauro Bitencourt de Souza[1] · Altamir Frederico Guidolin[1] · Sydney Antonio Frehner Kavalco[2] · Jefferson Luís Meirelles Coimbra[1]

[1] Department of Agronomy, Molecular Genetics and Breeding Institute, State University of Santa Catarina (UDESC) - Lages Campus, Lages, SC, Brazil.

[2] Agricultural Research and Rural Extension Company of Santa Catarina (EPAGRI), Chapecó, SC, Brazil

Corresponding author: Luan Tiago dos Santos Carbonari *carbonari.luan@gmail.com

Abstract

The study of genotype-by-environment interaction is essential in plant breeding, as stability and adaptability determine the success of genotype recommendations. Traditionally, numerous statistical methods and stability indices have been employed, yet they often face limitations when dealing with complex and nonlinear relationships. In this study, we aimed to apply Shapley values (SHAP), derived from game theory and widely used in machine learning models, as an alternative approach for interpreting genotype-by-environment interaction. The mean SHAP value proved to be strongly associated with adaptability, reflecting the average performance of genotypes, while the variability of SHAP values emerged as a reliable indicator of stability. Moreover, the method allowed an integrated assessment of environmental contributions, identifying favorable and unfavorable sites. Although it showed moderate agreement with traditional metrics, SHAP provided complementary insights, enhancing the understanding of genotype-by-environment interaction. Therefore, these metrics represent a promising tool to support decision-making in plant breeding programs, particularly for identifying superior genotypes across diverse environments.

Keywords phenotypic stability · adaptability · genotype recommendation · plant breeding

Introduction

Genotype-by-environment interaction (G×E) is a well-known phenomenon in plant breeding, characterized by the differential performance of genotypes when cultivated under varying environmental conditions. This interaction can complicate the selection and recommendation of superior genotypes, as a genotype that performs well in one environment may perform poorly in another (Falconer and Mackay 1996). Such behavior arises from the distinct responses of genotypes to environmental variations, which include both biotic factors (disease and pest pressure) and abiotic factors (temperature, precipitation, soil fertility, among others). Therefore, understanding and quantifying G×E interaction represents a key step in decision-making processes within breeding programs, influencing cultivar recommendation and strategies for cultivar release (Allard 1960).

Historically, the evaluation of genotype-by-environment interaction has been based on classical statistical methods that decompose the total observed variation into components attributable to genotype (G), environment (E), and the G×E interaction itself. These methods allow breeders to identify genotypes with desirable performance and adaptability across different environmental conditions. Among the classical approaches, those based on simple linear regression stand out, such as the method proposed by Finlay and Wilkinson (1963), who introduced the linear regression coefficient as a measure of genotype response to environmental gradients. Values close to 1 are interpreted as indicative of wide adaptation, values below 1 as adaptation to unfavorable environments, and values above 1 as adaptation to favorable environments.

Building on this concept, Eberhart and Russell (1966) incorporated, in addition to the regression coefficient (β_i), the variance of regression deviations ($\sigma^2_i d$) as a measure of stability. Ideal genotypes were defined as those with regression coefficient equal to one and variance of regression close to zero. Variance decomposition methods have also been widely used, such as that proposed by Wricke (1962), who estimated ecovalence (W_i) to measure each genotype's contribution to the interaction, assigning greater stability to those with lower ecovalence values. Similarly, Shukla (1972) presented a more precise estimator of the genotype-specific variance associated with the interaction, correcting for the bias caused by the number of environments and replications. In the same line of reasoning, Tai (1971) proposed the decomposition of the interaction into two parameters (α and λ), which quantify, respectively, the linear response and the consistency of genotype response to environmental variation.

In parallel, multivariate approaches have played a relevant role in the study of G×E interaction, as they allow for the exploration of complex data patterns. One of the most prominent is the AMMI model (Additive Main Effects and Multiplicative Interaction), described by Gauch (1988), which combines analysis of variance for the main effects of genotypes and environments with principal component analysis applied to the interaction. This method increases the explanatory

power and helps identify groups of genotypes with similar behavior. Another approach, the GGE Biplot method (Yan et al. 2000), focuses on the combined effects of genotype (G) and G×E interaction, allowing through graphical representation the identification of superior genotypes across “mega-environments” and the simultaneous evaluation of adaptability and stability.

Together, these methods form the methodological basis for stability and adaptability analysis in plant breeding, enabling the selection and recommendation of genotypes not only by their mean yield but also by their predictability and performance consistency across contrasting environments. Despite their widespread application, these methods have limitations, particularly their dependence on strict statistical assumptions, such as additivity of effects, normality and independence of errors, and homogeneity of variances. With the advancement of machine learning techniques, new approaches have emerged. Nonlinear models such as Random Forest and artificial neural networks have been applied because they can capture complex and high-order interactions among variables without the need to specify the functional form of the relationships in advance (Nascimento et al. 2013, Teodoro et al. 2015, Novielli et al. 2024). However, these models are often criticized as “black boxes”, making biological interpretation difficult, especially when the goal is to understand the relative contribution of genotype, environment, and their interaction to performance.

In this context, interpretability methods from data science offer opportunities to broaden the understanding of G×E interaction from a new perspective. Among these, the Shapley value, originally proposed by Lloyd Shapley (1953) in cooperative game theory, stands out. The Shapley value aims to fairly distribute the total gain obtained in a “game” among its participants, considering all possible orders of participation. When adapted to machine learning, this concept quantifies the average contribution of each variable (or model component) to prediction accuracy, considering all possible combinations.

Applied to the study of genotype-by-environment interaction, the Shapley value offers several advantages. It can be computed for groups of variables, allowing the isolation and quantification of the joint contribution of genotypic, environmental, and interaction information to the prediction of traits such as grain yield. This approach not only provides a relative measure of importance in percentage terms but also expresses the average impact of each group in the original response units (Lundberg and Lee 2017), which facilitates interpretation by plant breeders. Unlike traditional metrics that are restricted to decomposing observed variation, the Shapley value reflects the effective use of information by the predictive model, incorporating both linear and nonlinear effects (Lee et al. 2025, Dominic et al. 2024, Zhou et al. 2024). Another advantage of the Shapley value is its flexibility, since it can be applied to any predictive model, from classical linear models to complex algorithms. Therefore, it enables the study of genotype-by-environment interaction in

contexts where traditional methods lose explanatory power, such as when genotypic responses exhibit nonlinear behavior in relation to environmental conditions.

Thus, this study aims to propose and evaluate the use of the Shapley value as an alternative metric for interpreting genotype-by-environment interaction in plant breeding, exploring its potential to quantify the importance of genotypic, environmental, and interaction effects, thereby complementing the approaches traditionally used in stability and adaptability analysis.

Materials and Methods

Experimental setup and data used

This study utilized data from the *Value for Cultivation and Use* (VCU) trials of common bean (*Phaseolus vulgaris* L.) cultivars conducted in the state of Santa Catarina, Brazil, under the coordination of the Agricultural Research and Rural Extension Company of Santa Catarina (EPAGRI). The dataset originated from trials established across 12 distinct environments, resulting from combinations of different locations and growing seasons. The municipalities included Chapecó (first and second cropping seasons), Campos Novos, Ituporanga, Ponte Serrada, and Urussanga, with experiments conducted during the 2015/16 and 2016/17 agricultural years (Fig. 1).

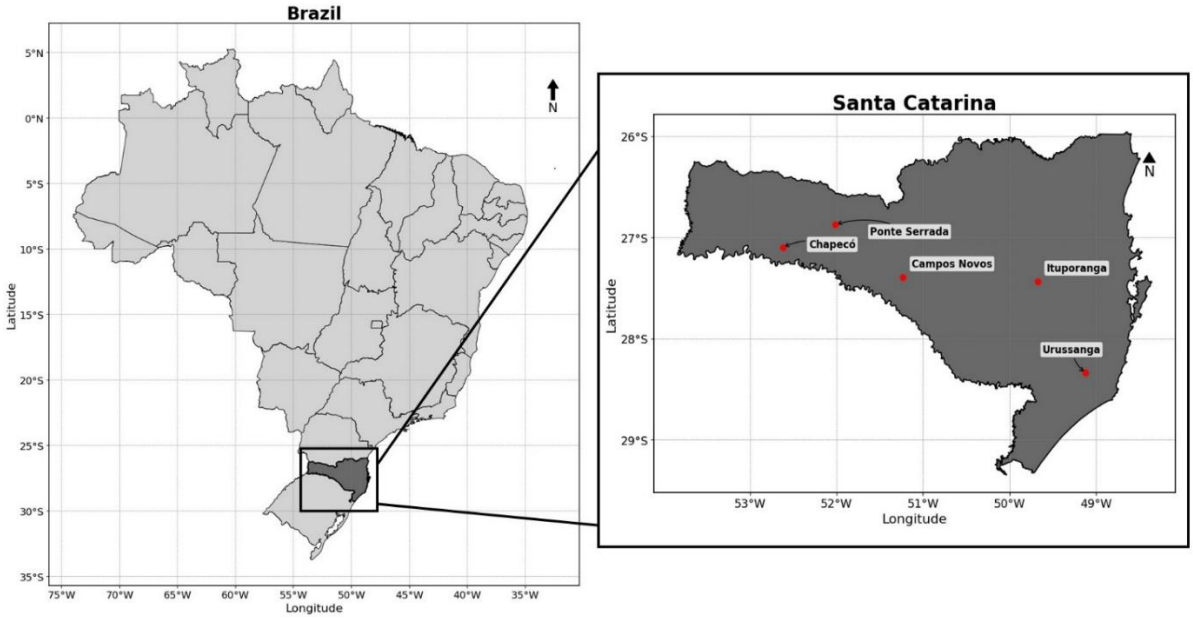


Fig. 1 Locations where the field experiments for testing the bean genotypes used in this study were conducted in the state of Santa Catarina, Brazil

The dataset comprised eight widely cultivated commercial genotypes that were common to all 12 environments: BRS Ametista, BRS Campeiro, BRS Esteio, IAC Imperador, IPR Campos Gerais, IPR Uirapuru, SCS 204 Predileto, and SCS 205 Riqueza. The experimental design used was a randomized complete block design (RCBD) with four replications per environment. Each

experimental plot consisted of four rows, each 4 m long and spaced 0.45 m apart, with a sowing density of approximately 15 seeds per linear meter.

Crop management practices followed the standard agronomic recommendations for common bean cultivation, including base and topdressing fertilization, seed treatment with fungicides and insecticides, and preventive pest and disease control. No supplemental irrigation was applied, so plants were fully exposed to the climatic conditions of each site and growing season. The trait evaluated in this study was grain yield, expressed in kilograms per hectare (kg ha^{-1}) and corrected to 13% grain moisture. Harvesting was performed in the useful area of each plot (the two central rows), excluding border rows and end plants to minimize border effects.

Data Analysis

The predictive model in this study was fitted using artificial neural networks, specifically through the MLPRegressor class of the scikit-learn package. The model was configured with two hidden layers containing 50 and 30 neurons, ReLU activation function, adam solver, a maximum of 1000 iterations, and a fixed random seed to ensure reproducibility. Training consisted of adjusting the synaptic weights based on grain yield data across different environments. This configuration allowed the model to capture complex and non-linear patterns commonly found in Value for Cultivation and Use (VCU) trials, in which the genotype × environment interaction plays a central role. The use of neural networks in this context strengthens the model's generalization capacity and supports the application of SHAP values as an interpretative tool.

In general, SHAP values (SHapley Additive exPlanations) are grounded in cooperative game theory and quantify the contribution of each explanatory variable to the prediction of a machine learning model. Each SHAP value indicates how much a variable contributes to increasing or decreasing the prediction for a specific observation. Positive values indicate an increase, negative values indicate a decrease, and higher absolute values reflect a greater impact of the variable. The computation of SHAP values considers all possible combinations of inclusion and exclusion of the model's components, ensuring a fair attribution of contributions. Formally, for a prediction f(x)and a set Sof features, the SHAP value for feature jin observation xis given by:

$$SHAP_j(x) = \sum_{F \subseteq S\{j\}} \frac{|F|!\,(|S|-|F|-1)!}{|S|!}[f_x(F \cup \{j\}) - fx(F)]$$

where $|F|!$and $(|S|-|F|-1)!$are weighting factors based on the number of feature permutations, and $f_x(F \cup \{j\})$and $f_x(F)$are the model predictions with and without feature j, respectively (Lundberg and Lee, 2017).

In this study, the environment and genotype vectors were encoded using *one-hot encoding*, as well as their explicit interaction (genotype × environment), allowing the model to directly capture these effects. SHAP values were then calculated for each level of the effects, enabling the

decomposition of the prediction for each observation $i$into the sum of the contributions of individual features:

$$\widehat{y_i} = \phi_0 + \sum_{j=1}^{p} \phi_{ij},$$

where $\phi_0$is the average model prediction and $\phi_{ij}$is the SHAP value of feature $j$for observation i. To interpret the effect contributions, the dummy variables were grouped into three sets: *i)* G: indices of dummy variables associated with genotypes; *ii)* E: indices of dummy variables associated with environments; and *iii)* $G \times E$: indices of dummy variables associated with the genotype × environment interaction. Thus, for each observation i, the contributions were computed as follows:

$SHAP_{G,i} = \sum_{j \in G} \phi_{ij}$ (genotype effect)

$SHAP_{E,i} = \sum_{j \in E} \phi_{ij}$ (environment effect)

$SHAP_{GxE,i} = \sum_{j \in GxE} \phi_{ij}$ (genotype × environment interaction)

Considering that the interaction represents the joint and inseparable effect of genotype and environment, a proportional redistribution of this contribution was adopted. Consequently, the SHAP-based analysis enabled the decomposition of the predictive effect of the non-linear model into interpretable components, explicitly reflecting the effects of genotype, environment, and their interaction, thereby contributing to a deeper understanding of yield variability under different conditions.

The performance of this approach was compared with traditional methods such as that proposed by Eberhart and Russell (1966), which relies on a linear regression analysis of the mean performance of each genotype across environments against an environmental index. The yield data were organized according to a randomized complete block design with replications across multiple environments (locations and/or years). The environmental index (I_j) was obtained as the difference between the mean of all genotypes in environment $j$and the overall experimental mean, representing the relative productivity potential of each environment. Then, for each genotype i, the following statistical model was fitted:

$$Y_{ij} = \mu_i + \beta_i X_j + \delta_{ij},$$

where $Y_{ij}$is the yield of genotype $i$in environment j, $\mu_i$is the mean yield of genotype i, $\beta_i$is the regression coefficient indicating adaptability, and $\delta_{ij}$is the deviation from regression (unexplained variation). The parameter $\beta_i$is interpreted as a measure of adaptability: values close to 1 indicate broad adaptation, $\beta_i > 1$indicate higher responsiveness to favorable environments, and $\beta_i <$ 1suggest better performance under restrictive conditions. Stability is assessed by the variance of regression deviations (S_{di}^2), where values close to zero indicate greater phenotypic stability. In practice, genotypes considered ideal according to this method are those with $\beta_i \approx 1$, S_{di}^2not significantly different from zero, and mean yield above the overall mean. This approach enables the

separation of adaptability (mean response to the environmental gradient) and stability (consistency of response), providing robust information for cultivar recommendation and selection.

Another method used for comparison was the AMMI (*Additive Main Effects and Multiplicative Interaction*) model, as described by Gauch (2013):

$$Y_{ijk} = \mu + e_j + be_j + g_i + \sum_{k=1}^{n} \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + \mathcal{E}_{ijk},$$

where $Y_{ijk}$is the observation of the i-th genotype (i = 1, 2, ..., 8), in the j-th environment (j = 1, 2, ..., 12), and in the k-th block (k = 1, 2, 3, 4); $\mu$is the overall mean; $e_j$is the main effect of environment; $b_{ej}$is the block effect within environment j; $g_i$is the main effect of genotype; λ_k, γ_{ik}, and $\alpha_{jk}$are the singular value decomposition (SVD) terms of the genotype × environment interaction matrix; ρ_{ij}represents the residual noise related to the interaction; and $\varepsilon_{ijk}$is the experimental error. The stopping point in selecting the AMMI model (PCA1, PCA2, ..., PCAn) was determined based on the significance of successive F-tests for the interaction terms, with the remaining non-significant components grouped into complementary portions of the G×E sum of squares. The AMMI Stability Value (ASV) was calculated as:

$$ASV = \{((SQ_{PC1}/SQ_{PC2}).(PC1))^2 + PC2^2\}^{1/2},$$

where $ASV$is the stability value, $SS_{PC1}$and $SS_{PC2}$are the sums of squares for the first and second interaction principal components, and $PC1$and $PC2$are the respective genotypic scores.

The comparison among SHAP, AMMI, and Eberhart & Russell methods was based on measures of association and concordance. First, the Spearman correlation coefficient (ρ) was calculated between the genotypic scores obtained by each method to assess the monotonic association between adaptability and stability metrics. Then, rank concordance was evaluated using Kendall's coefficient (τ), which quantifies the similarity between the genotypic rankings (Kendall and Smith, 1939). Finally, the overlap between the top-performing genotypes was assessed using the Jaccard index (1901), considering the five best genotypes identified by each method, as defined by their respective criteria (higher mean yield, lower instability, or greater positive SHAP contribution). The index was calculated as $|A \cap B| / |A \cup B|$, where $A$and Brepresent the sets of genotypes selected by the compared methods, thus quantifying the similarity between their recommendations. All analyses were performed in Python. The *pandas* and *numpy* libraries were essential for data handling and organization, while *scikit-learn* was used for preprocessing (via *ColumnTransformer* and *OneHotEncoder*) and for model fitting with artificial neural networks (*MLPRegressor*). Model interpretability was achieved through the *SHAP* library, which quantified the relative contributions of genotypes, environments, and their interactions. The *scipy.stats* library supported the computation of correlations and regressions, while *matplotlib* was used for graphical visualization. This integrated set of tools enabled a complete workflow from data preparation and modeling to detailed interpretation of the results.

Results

Initially, the graphical visualization of SHAP values enables an integrated understanding of the behavior and variation in each common bean genotype's contribution to predicted grain yield using the neural network algorithm (Fig. 2). Each mean point represents the expected effect of a genotype on prediction, while the error bars indicate the standard deviations, reflecting how much this contribution varied. Thus, positive SHAP values indicate that the genotype contributed, on average, to increasing grain yield in the model (indicating superior performance), whereas negative values denote a contribution toward reducing the estimates, reflecting inferior performance under the evaluated environmental conditions. Therefore, this metric allows quantifying the relative importance of each genotype while explicitly accounting for environmental influence and genotype-by-environment interaction.

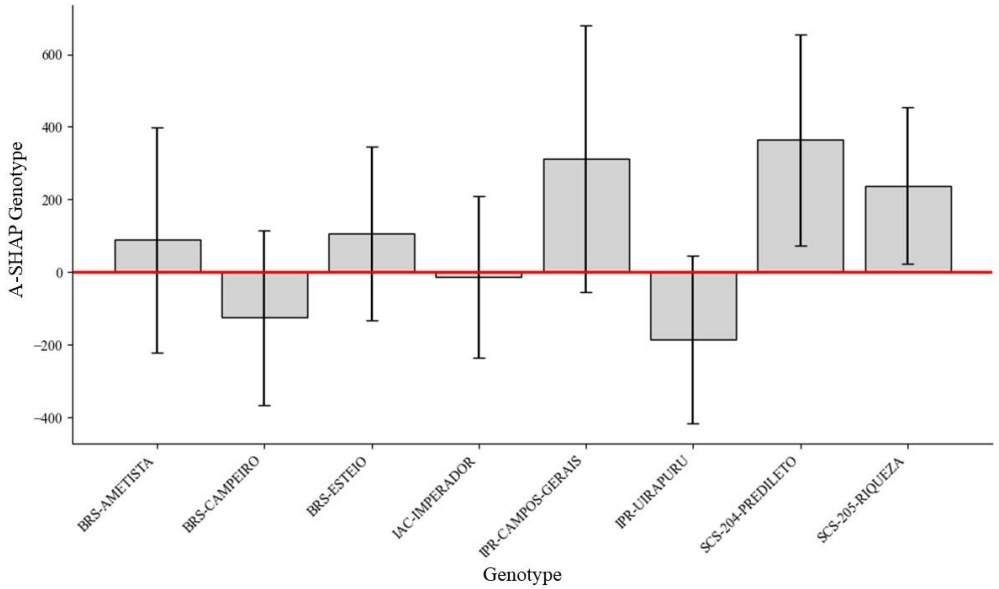


Fig. 2 Mean A-SHAP values for different bean genotypes, with error bars representing standard deviations, indicating the contribution of each genotype to the predicted grain yield

As an example, the genotypes IPR Campos Gerais, SCS 204 Predileto, and SCS 205 Riqueza stood out by showing higher and positive mean SHAP values, suggesting that, regardless of specific environmental conditions, these genotypes tended to maintain higher yield estimates. This behavior can be interpreted as evidence of broad adaptability, since these genotypes consistently contribute favorably to grain yield. In contrast, genotypes such as BRS Campeiro and IPR Uirapuru showed negative mean SHAP values, indicating an unfavorable contribution to yield, which suggests restricted adaptability or a limited capacity to exploit environmental resources. Genotypes with SHAP values near zero, such as BRS Ametista, exhibited intermediate or neutral behavior in terms of average contribution, which may reflect strong dependence on specific environmental conditions.

In the context of plant breeding, phenotypic adaptability can be directly related to the mean SHAP value. Positive and consistent SHAP means across environments indicate that the genotype can perform well under different conditions, characterizing broad adaptability. Negative values indicate poor performance in most scenarios, while values close to zero suggest a neutral effect, meaning that the genotype does not stand out compared to the population average for the trait. Genotypic stability, in turn, is related to the variation in contributions, represented by the standard deviation of SHAP values (S-SHAP). Low deviations indicate consistent performance across environments (that is, the SHAP effect is little influenced by environmental variation), while high deviations reflect greater environmental sensitivity. For instance, IPR Campos Gerais combined high adaptability (positive SHAP mean) with moderate stability (low deviation), while BRS Ametista showed null adaptability with low stability (high deviation), revealing a strong dependence on environmental conditions for yield expression.

Considering the results from the environmental perspective, the contribution of each environment to grain yield prediction is demonstrated (Fig. 3). The mean E-SHAP values for each environment, along with standard deviation bars, allow the identification of environments that were favorable, unfavorable, or neutral for common bean cultivation. Environments such as Ponte Serrada (16-17), Chapecó (16-17), and Campos Novos (16-17) exhibited positive values and were therefore considered favorable, since local conditions enhanced yield estimates. In contrast, environments such as Campos Novos (15-16), Urussanga (15-16), and Ponte Serrada (15-16) presented negative values and were classified as unfavorable, as they reduced the predicted productive potential. Environments with values close to zero, such as Chapecó (16-17) and Canoinhas (15-16), were considered buffered or neutral, showing no significant influence on yield prediction, suggesting that in these cases, yield depended more on genotypic contributions than on environmental effects.

The interpretation of environments follows the same rationale as for genotypes. Positive E-SHAP means indicate favorable environmental conditions for yield expression, while negative means suggest environmental constraints, such as biotic or abiotic stresses or inadequate management. The variability associated with each environment is directly related to the standard deviation of E-SHAP values: low deviations indicate that the environment was consistent across genotypes, whereas high deviations suggest more pronounced genotype-by-environment responses. For example, Ponte Serrada (16-17) not only showed positive mean values but also low deviations, indicating a favorable and stable environment, suitable for identifying high-yielding genotypes. In contrast, Campos Novos (15-16) had negative mean values and high deviations, suggesting a restrictive environment in which few genotypes performed well, useful for identifying superior genotypes under stress-prone conditions.

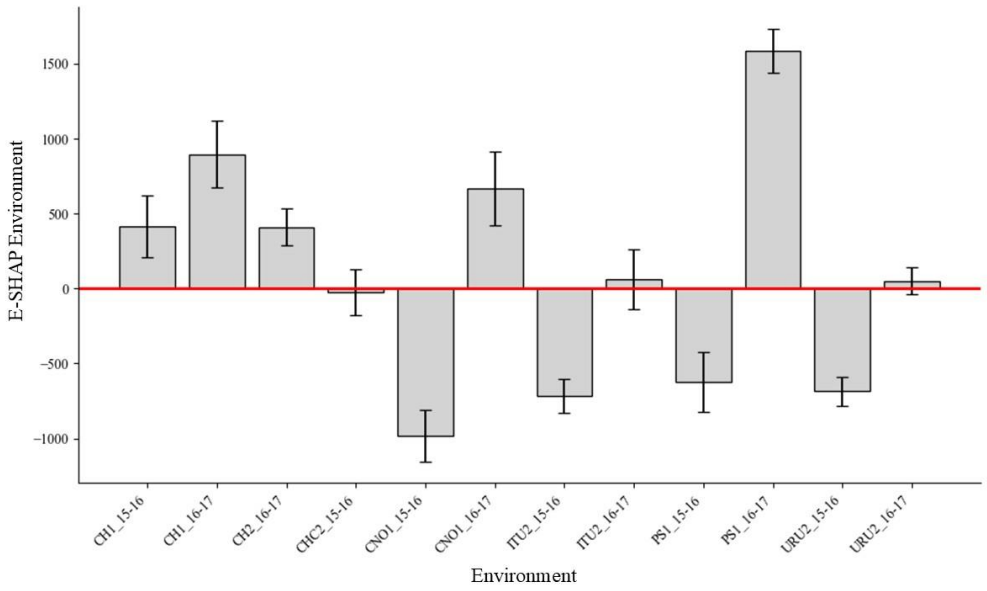


Fig. 3 Mean E-SHAP values for different bean growing environments, with error bars representing standard deviations, highlighting the contribution of each environment to the predicted grain yield by the neural network model

This SHAP-based approach offers a significant advantage because it enables simultaneous interpretation of adaptability and stability for genotypes, along with information about environmental contributions and variability. Unlike traditional methods that often separate main and interaction effects, SHAP analysis maintains an integrated view, highlighting not only which genotypes or environments stand out but also how their contributions vary. Accordingly, the reaction norms of SHAP values for each genotype across environments (Fig. 4) provide a comprehensive interpretation of both genotype performance and environmental influence. These results are consistent with those observed for A-SHAP means, where the best-performing genotypes across environments were IPR Campos Gerais and SCS 204 Predileto, while BRS Campeiro and IPR Uirapuru showed inferior performance.

The values of stability, adaptability, and grain yield for the eight common bean genotypes, evaluated using SHAP, Eberhart and Russell, and AMMI methods, are presented in Table 1. Stability was measured using the standard deviation of SHAP values (S-SHAP), the variance of regression deviations (S^2di), and the AMMI Stability Value (ASV). Lower values of S-SHAP, S^2di, and ASV indicate higher stability (that is, smaller variation in yield across environments). The genotypes SCS 205 Riqueza and IAC Imperador showed the lowest S^2di (34,471 and 21,704, respectively) and ASV (380.50 and 96.74), indicating higher stability, while IPR Campos Gerais and BRS Ametista showed lower stability according to all three methods.

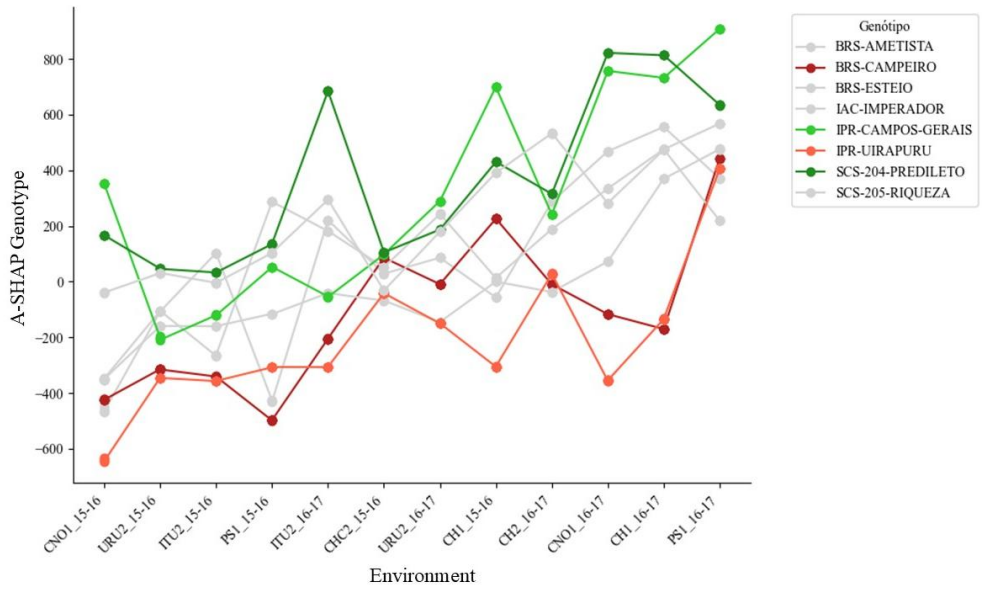


Fig. 4 Variation of A-SHAP values for different bean genotypes across growing environments, highlighting the contribution of each genotype to the predicted grain yield using neural networks

Adaptability was assessed using the mean A-SHAP value, representing the average genotypic effect across environments for grain yield, and the regression coefficient (βi), indicating the genotype's response to environmental improvement. Positive SHAP values and βi close to 1 suggest broad adaptability. In this context, SCS 204 Predileto (SHAP = 364.43; βi = 1.00) and IPR Campos Gerais (SHAP = 312.89; βi = 1.20) exhibited high adaptability, whereas IPR Uirapuru (SHAP = -185.64; βi = 0.92) and BRS Campeiro (SHAP = -124.62; βi = 0.93) showed lower adaptability.

Mean grain yield ranged from 2,183 kg ha^{-1} (IPR Uirapuru) to 2,732 kg ha^{-1} (SCS 204 Predileto), with an overall mean of 2,447 kg ha^{-1}. Genotypes with the highest yields did not always coincide with the most stable ones. For instance, SCS 204 Predileto exhibited the highest yield and high adaptability but intermediate stability, whereas IAC Imperador combined stability with yield (2,301 kg ha^{-1}). These results highlight that the selection of superior genotypes depends on the breeding objective. For broad recommendation, genotypes with positive SHAP values and βi close to 1 are preferable, along with moderate stability. For specific environments, genotypes with greater adaptation to favorable or unfavorable conditions may be prioritized. The integration of SHAP, Eberhart and Russell, and AMMI methods confirms the robustness of SHAP metrics for evaluating genotype-by-environment interaction in plant breeding programs.

Table 1 Stability, adaptability, and grain yield values obtained for eight bean genotypes, including SHAP, Eberhart and Russell, and AMMI metrics

Genotype	Stability			Adaptability		Yield
	S-SHAP	S^2d_i	*ASV*	*A-SHAP*	*βi*	
BRS Ametista	310.54	177057	1186.30	88.55	1.03	2427
BRS Campeiro	240.17	161524	1094.79	-124.62	0.93	2188
BRS Esteio	238.64	87774	231.41	105.98	0.98	2525
IAC Imperador	222.51	21704	96.74	-13.48	0.96	2301
IPR Campos Gerais	367.06	178573	1443.00	312.89	1.20	2612
IPR Uirapuru	231.27	140208	1152.98	-185.64	0.92	2183
SCS 204 Predileto	290.63	122775	997.19	364.43	1.00	2732
SCS 205 Riqueza	215.33	34471	380.50	237.30	0.98	2609
Média	264.52	115511	822.86	98.18	1.00	2447

Stability was estimated by the standard deviation of SHAP values (S-SHAP), by the Eberhart and Russell method (S^2di), and by the AMMI stability value (ASV). Adaptability was estimated by the mean SHAP value (A-SHAP) and the regression coefficient (βi), while productive performance was expressed as the mean grain yield (kg ha^{-1}) obtained across the evaluated environments.

Table 2 summarizes the comparative performance of the different methods used to evaluate stability and adaptability for grain yield in common bean genotypes through three complementary metrics: Pearson's correlation coefficient, Jaccard similarity index, and Kendall's concordance coefficient, which together quantify the consistency among approaches for classifying stability, adaptability, and mean yield. Regarding stability, the analyses included the SHAP standard deviation (S-SHAP), the classical Eberhart and Russell stability parameter (S^2di), and the AMMI Stability Value (ASV). Pearson's correlation analysis revealed strong positive associations among these methods, with coefficients of 0.70 between S-SHAP and S^2di and 0.71 between S-SHAP and ASV, indicating substantial agreement in genotype stability ranking. Conversely, the correlation between S-SHAP and mean grain yield was moderate (0.45), reinforcing that stability estimated through SHAP values is only partially related to productivity, which is consistent with previous studies reporting that high yield and phenotypic stability are not necessarily associated.

The Jaccard similarity index, which measures overlap between classified sets, showed intermediate agreement among stability methods. Between S-SHAP and S^2di, the index was 0.45, while between S-SHAP and ASV, it was slightly lower (0.42), suggesting partial agreement in identifying stable genotypes. Conversely, S-SHAP versus yield showed a higher index (0.66), implying that despite weak linear correlation, there was some similarity in genotype classification based on stability and yield. Kendall's concordance coefficient, which measures agreement among rankings, yielded results consistent with Pearson correlations. The concordance between S-SHAP

and S²di was 0.71 (good agreement), between S-SHAP and ASV 0.57 (moderate agreement), and between S-SHAP and yield 0.21 (low agreement), reinforcing that stability does not necessarily imply higher yield.

Table 2 Pearson correlation coefficients, Jaccard similarity indices, and Kendall concordance coefficients obtained for the comparisons among stability, adaptability, and grain yield of bean genotypes

Parameter	Comparison	Estimates		
		Correlation (ρ)	Jaccard (J)	Kendall (τ)
Stability*	*S-SHAP and* S^2d_i	0.70	0.45	0.71
	S-SHAP and ASV	0.71	0.42	0.57
	S-SHAP and Grain yield	0.45	0.66	0.21
Adaptability**	*A-SHAP and βi*	0.65	1.00	0.71
	A-SHAP and Grain yield	0.96	1.00	1.00

*Stability was represented by the standard deviation of SHAP values (S-SHAP), by the Eberhart and Russell method (S²di), and by the AMMI stability value (ASV). **Adaptability corresponds to the mean SHAP value (A-SHAP) and the regression coefficient (βi), while grain yield (kg ha^{-1}) was obtained across the evaluated environments.

For adaptability, assessed through mean SHAP and regression coefficient (βi), the metrics indicated high consistency. Pearson's correlation between SHAP and βi was 0.65, while correlation with yield was extremely high (0.96), showing that SHAP-based adaptability closely reflects mean yield performance. The Jaccard and Kendall indices showed perfect agreement (1.0) for SHAP versus βi and SHAP versus yield, indicating complete consistency in genotype ranking and classification, confirming that more adaptable genotypes also tend to yield more. In summary, these results show that while stability metrics display moderate agreement with traditional indices, adaptability measured by SHAP is strongly associated with both βi and yield, supporting SHAP values as a reliable and integrative tool for evaluating genotype performance in common bean breeding programs.

Discussion

Interpretation of SHAP values

The concept of genetic correlation provides an important theoretical foundation for understanding genotype-by-environment interaction. When the same trait is evaluated under distinct environmental conditions, it should not be interpreted as a single variable, but rather as two potentially different expressions of the same biological phenomenon. This occurs because the physiological mechanisms underlying its expression may vary according to the environment, meaning that different sets of genes may be involved in determining performance (Falconer 1952).

Thus, when the genetic correlation between two environments is high, it can be inferred that, to a large extent, the same genes are acting in both. Conversely, when this correlation is low or close to zero, it suggests that different genes contribute to trait expression in each environment, representing a classical case of genotype-by-environment interaction. This interaction therefore reflects the differential variation of phenotypic values of genotypes across distinct environmental contexts and is inseparable from environmental variance (Falconer and Mackay 1996). This condition implies that the performance of a genotype is not absolute, but relative to the environment in which it is evaluated, meaning that the best genotype in one location may not be the best in another.

This conceptual framework supports the importance of metrics capable of simultaneously quantifying the contribution of genotypes, environments, and their interaction for complex traits such as grain yield. In this context, Shapley values represent a promising alternative. Derived from cooperative game theory, SHAP allows the decomposition of predictions from machine learning models into individual contributions, enabling direct estimation of how much each genotype or environment adds to or reduces the prediction. Therefore, the mean SHAP value for genotypes can be interpreted as a measure of adaptability (A-SHAP), while its variability (S-SHAP) reflects the stability of genotypes across different environments, fully consistent with the logic of classical analyses of genotype-by-environment interaction (Allard and Bradshaw 1964).

The present approach offers significant advantages compared with traditional methods. While the method of Eberhart and Russell (1966) relies on assumptions of linearity and homogeneity of variances, and the AMMI model requires multivariate decomposition of effects (Gauch 1988), SHAP provides an intuitive and comparable synthesis of the expected contribution of each component (genotype, environment, and interaction) to grain yield. Furthermore, by preserving environmental variability in the modeling process, SHAP reinforces the active role of environments in differentiating genotypes, aligning with the perspective that genotype-by-environment interaction should not be regarded merely as a source of noise, but as a fundamental component for cultivar recommendation.

In the results presented in this study, bean genotypes such as BRS Campos Gerais, SCS 204 Predileto, and SCS 205 Riqueza stood out by showing positive and relatively consistent mean A-SHAP values, suggesting broad adaptability and the ability to maintain high performance under diverse environmental conditions. This finding supports the expectation that genotypes with stable and above-average performance are promising candidates for wide recommendation, especially for breeding programs aiming to serve multiple production regions.

Conversely, genotypes such as BRS Campeiro and IPR Uirapuru exhibited negative mean A-SHAP values, indicating an unfavorable contribution to grain yield. This evidence points to

restricted adaptability or inherent physiological limitations of the genotype in exploiting environmental resources. Although such genotypes are less promising for broad recommendation, they can be valuable in specific environments or used in crosses to broaden the genetic base, providing alleles of interest for complementary traits (Allard 1960).

Another relevant aspect was observed in the genotype BRS Ametista, which showed mean values close to zero accompanied by higher variability. This suggests that the performance of this genotype depends strongly on specific environmental conditions, representing a typical case of interaction. Such information is extremely relevant for breeding programs, as it allows the identification of genotypes with potential for restricted recommendation, targeted to regions with particular conditions that favor their performance.

The main contribution of Shapley values in this context is to provide a metric that is both intuitive and comparable across genotypes. While the Eberhart and Russell method depends on assumptions of linearity and homogeneity of variances, and the AMMI model requires multivariate decomposition of effects, SHAP summarizes the expected contribution of each genotype to the yield predicted by the neural network model. This integrative capability, while accounting for environmental variability, represents a notable advantage over classical methods.

Figure 3 shifts the focus of analysis from genotypes to environments, allowing the identification of which locations and years favored or limited the productive performance of common bean. Some environments (Ponte Serrada 16-17, Chapecó 16-17, and Campos Novos 16-17) exhibited positive E-SHAP values with low deviations, characterizing them as favorable and consistent environments for the expression of the genotypic potential for grain yield. In contrast, some environments (Campos Novos 15-16, Urussanga 15-16, and Ponte Serrada 15-16) showed negative values, indicating unfavorable growing conditions. This interpretation aligns with the concept of discriminating environments discussed by Yan and Tinker (2006) in the context of GGE Biplot analysis, in which locations that capture meaningful fractions of variation contribute to the identification of superior genotypes. However, SHAP values provide an additional advantage, offering the possibility of directly quantifying the expected effect of each environment on the prediction, rather than relying on indirect relationships through principal component scores.

Another relevant aspect is the variability of E-SHAP values among different environments. Environments with low standard deviation (Ponte Serrada 16-17) indicate homogeneous conditions where genotypes responded similarly. In contrast, environments with high deviations (Campos Novos 15-16) reflect greater environmental variability, evidencing strong genotype-by-environment interaction. This distinction is highly relevant for breeding programs, as environments with limited variation, although generally unfavorable, are essential for identifying genotypes with superior performance under restrictive conditions.

Thus, analysis based on Shapley values allows not only the identification of more productive environments but also the assessment of their contribution to the selection process. This approach reinforces the notion that the environment should not be viewed merely as a source of experimental noise, but as an active component of the interaction, whose characterization is fundamental for cultivar recommendation.

Comparison between SHAP and traditional methods

The comparison between methods allows the assessment of the consistency and complementarity of the metrics. For stability, the correlations among the metrics (S-SHAP, S^2di, and ASV) are moderate, indicating positive relationships. This finding suggests that, although the methods capture the essence of stability, there are distinctions in how each metric reflects genotypic response. Stability measured by S-SHAP evaluates the variability of each genotype's contributions within the predictive model (Dominic et al. 2022), whereas S^2di is based on the variance of regression deviations, and ASV relies on multivariate decomposition using the AMMI model. Therefore, it is not surprising that the agreement is not perfect, as these are mathematically distinct approaches, which consequently translate into different concepts of stability for each method.

The Jaccard index and Kendall's coefficient reinforce this interpretation, showing that the ranking of genotypes in terms of stability coincides partially across methods. This divergence, far from being a limitation, can be seen as complementarity, as different metrics provide multiple perspectives on the same phenomenon (Danilevicz et al. 2025). For adaptability, the results were consistent. Correlations among mean A-SHAP values, the regression coefficient (βi), and mean grain yield were high, particularly for A-SHAP versus yield (0.96). Furthermore, the Jaccard and Kendall indices indicated maximum agreement, showing that the rankings obtained by the methods were identical (A-SHAP, βi, and mean grain yield). These results suggest that A-SHAP is efficient in capturing adaptability responses, presenting equivalence with a classical method while also offering greater flexibility for applying predictive models in the study of genotype-by-environment interaction.

This convergence between methods (A-SHAP and βi) can be interpreted by the very definition of adaptability as the ability of genotypes to respond to different environmental conditions (Vencovsky and Barriga 1992). In this context, the regression coefficient quantifies this response relative to a linear environmental gradient, while SHAP estimates the contribution of genotypes to environments within a non-linear model. The fact that both metrics converge, despite arising from different assumptions, reinforces the robustness of Shapley values as an alternative method for studying adaptability in plant breeding.

To date, no studies have been found in the literature applying SHAP values directly for the evaluation of stability and adaptability in plant breeding. Generally, reported applications focus on the interpretation of predictive models, highlighting the relative contribution of each trait or variable in predicting a phenotype (Lee et al. 2025). A similar approach was observed in a study that integrated the SHAP explanatory module, allowing the ranking of variable importance and individualized interpretation of predictions (Yu et al. 2025). In these cases, the technique demonstrated potential to reveal key genotypic and environmental factors determining yield predictions, even without specifically targeting genotype-by-environment interaction analysis.

Thus, it is evident that the literature on stability and adaptability is still dominated by classical methodologies, each with its own statistical foundation and specific applications. Recent studies have emphasized the need to carefully compare different methods, recognizing that each approach may highlight distinct aspects of genotype-by-environment interaction (Yao et al. 2025, Nascimento et al. 2013, Teodoro et al. 2015). In this context, introducing Shapley values as an alternative metric expands the analytical repertoire, enabling an integrated interpretation that complements traditional methods while leveraging the flexibility of machine learning models.

Implications for Plant Breeding

The adoption of interpretable metrics, such as SHAP, in plant breeding has direct and innovative implications for decision-making in selection programs and cultivar recommendations. Traditionally, the analysis of genotype-by-environment interaction has been conducted using linear biometric models, such as regressions, indices, and multivariate approaches (Finlay and Wilkinson 1963, Eberhart and Russell 1966, Gauch 2013). While these methods are established as robust tools for evaluating adaptability and stability, they still have limitations in capturing complex and non-linear relationships, particularly in crops where genotype responses to environments are influenced by multiple biotic and abiotic factors. In this context, the use of Shapley values emerges as a complementary methodological alternative, capable of introducing innovation and facilitating the interpretation of complex models such as neural networks.

First, the main contribution of this approach to plant breeding is the ability to simultaneously evaluate the contributions of genotypes and environments, providing an integrated view of genotype-by-environment interaction. This allows the identification not only of which genotypes stand out in terms of mean yield or stability but also the understanding of the environmental conditions in which their superiority is expressed. This approach aligns with contemporary breeding goals, which aim not only to identify superior cultivars but also to understand the mechanisms underlying their performance across environments (Cooper et al. 2021).

Second, SHAP expands the generalization capacity of analyses. Being derived from artificial intelligence models, such as the neural networks used in this study, it incorporates non-linear relationships and higher-order interactions among factors that are rarely captured by traditional statistical models. This aspect is particularly relevant for common bean, a crop sensitive to edaphoclimatic variations. Consequently, SHAP enables the identification of complex adaptation patterns, supporting strategies for cultivar recommendation.

Another important point refers to the flexibility of applying this explanatory technique to meet different objectives within a breeding program. In situations targeting broad recommendations, genotypes with positive mean values and low variability are preferred, as observed in cultivars SCS-204 Predileto and SCS-205 Riqueza, which exhibited consistent performance across multiple environments. In contrast, for specific recommendations, genotypes with intermediate contributions and higher variability, such as BRS Ametista, become attractive, particularly when associated with environments previously characterized as favorable. This flexibility reflects a central feature of modern breeding, aligning selection strategies with the specific demands of environments and cropping systems (Ceccarelli 2015).

Method integration also strengthens result interpretation. In this study, the agreement of SHAP with classical adaptability metrics and its moderate consistency with stability metrics indicates that its use should not be seen as a replacement but as a methodological advancement that combines breeding with artificial intelligence. In other words, this approach serves as a complementary tool, providing enhanced interpretive power in genotype-by-environment interaction analyses and enriching the decision-making process.

Overall, the results demonstrate that Shapley values constitute a promising metric for integrated analysis of adaptability and stability in breeding programs. The simultaneous interpretation for both genotypes and environments offers a comprehensive and intuitive understanding of the interaction phenomenon. Practically, A-SHAP showed strong association with adaptability and mean grain yield, validating its use. Although stability presents moderate agreement with traditional methods, this difference should be understood as reflecting distinct concepts and mathematical foundations. The joint use of metrics can thus enrich recommendations and increase confidence in decision-making.

By enabling integration between predictive models based on artificial intelligence and interpretable statistical metrics, Shapley values represent a significant advance for plant breeding and can be applied to any crop. Their adoption can accelerate the identification of genotypes adapted to diverse scenarios, improve the accuracy of cultivar recommendations, and ultimately enhance the efficiency of breeding programs. In summary, the main contribution of this work was to

demonstrate a complement and extension of traditional approaches, offering a flexible alternative for studying genotype-by-environment interaction in plant breeding.

Conclusions

Shapley values are effective for interpreting genotype-by-environment interaction, allowing the quantification of genotype adaptability and stability in a way that complements traditional metrics. Adaptability (A-SHAP) derived from this metric shows strong association with the regression coefficient and yield, enabling the identification of genotypes that are broadly adaptable or restricted to specific conditions. The variability of the values allows for the assessment of stability (S-SHAP) in response to environmental fluctuations, despite moderate agreement with classical indices. Shapley values also enable the evaluation of environmental contributions (E-SHAP), distinguishing favorable and unfavorable locations. Therefore, indices based on Shapley values represent a promising alternative to complement conventional approaches, with the added advantage of integrating with artificial intelligence-based models, capturing non-linear relationships, and contributing to greater accuracy and efficiency in plant breeding programs.

Acknowledgments

We thank the Foundation for Research and Innovation Support of the State of Santa Catarina (FAPESC), Coordination for the Improvement of Higher Education Personnel (CAPES) and the University of the State of Santa Catarina (UDESC) for granting scholarships and financial assistance in carrying out research projects.

Data availability

All data related to the analyses performed are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that there is no conflict of interest.

Funding

This research was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES) and the University of the State of Santa Catarina (UDESC).

Author contributions

All authors are listed in alphabetical order and contribute equally. Luan Tiago dos Santos Carbonari: Conceptualization, Methodology, Formal analysis, Writing-Original Draft. Carlos Zacarias Joaquim Júnior and Mauro Bitencourt de Souza: Conceptualization, Methodology and Formal analysis. Altamir Frederico Guidolin, Sydney Antonio Frehner Kavalco and Jefferson Luís Meirelles Coimbra: Data Curation, Writing-Review and Editing, Funding acquisition. All authors read and approved the final manuscript.

Referências

Allard, R. W. (1960). Principles of plant breeding. New York, NY: John Wiley & Sons.

Allard, R. W., & Bradshaw, A. D. (1964). Implications of Genotype-Environmental Interactions in Applied Plant Breeding. Crop Science, 4(5), 503-508.

Ceccarelli, S. (2015). Efficiency of plant breeding. *Crop Science, 55*(1), 87-97. https://doi.org/10.2135/cropsci2014.02.0158

Cooper, M., Technow, F., Messina, C., Gho, C., & Totir, L. R. (2021). Use of crop growth models with whole-genome prediction: Application to a maize multienvironment trial. *Crop Science, 61*(1), 98-117. https://doi.org/10.1002/csc2.20345

Danilevicz, M.F., Upadhyaya, S.R., Batley, J., Bennamoun, M., Bayer, P.E. and Edwards, D. (2025), Understanding plant phenotypes in crop breeding through explainable AI. Plant Biotechnol. J. https://doi.org/10.1111/pbi.70208

Dominic, N., Cenggoro, T. W., Budiarto, A., Pardamean, B., et al. (2022, August 15). Deep polygenic neural network for predicting and identifying yield-associated genes in Indonesian rice accessions. Scientific Reports, 12, Article 13823. https://doi.org/10.1038/s41598-022-16075-9

Eberhart, S. A., & Russell, W. A. (1966). Stability parameters for comparing varieties. *Crop Science, 6*(1), 36-40. https://doi.org/10.2135/cropsci1966.0011183X000600010011x

Falconer, D. S., & Mackay, T. F. C. (1996). Introduction to quantitative genetics (4th ed.). Longman.

Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research, 14*(6), 742-754. https://doi.org/10.1071/AR9630742

Gauch, H. G. (1988). Model selection and validation for yield trials with interaction. Biometrics, 44(3), 705-715.

Gauch, H. G. A simple protocol for AMMI analysis of yield trials. Crop Science, v. 53, n.5, p. 1860-1869, 2013. https://doi.org/10.2135/ cropsci2013.04.0241.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull Soc Vaudoise Sci Nat, 37, 547-579.

Kassem, A. Harnessing Artificial Intelligence and Machine Learning for Identifying Quantitative Trait Loci (QTL) Associated with Seed Quality Traits in Crops. Plants, 14(11), 1727. https://doi.org/10.3390/plants14111727

Kendall, M. G., & Smith, B. B. (1939). The problem of m rankings. The annals of mathematical statistics, 10(3), 275-287.

Lee, S. Y., Lee, H., Han, J., Jeung, J.-U., & Mo, Y. (2025, June 10). Integrating genotype-by-environment interaction and machine learning to evaluate yield-related traits in Korean rice cultivars. Journal of Crop Science and Biotechnology. https://doi.org/10.1007/s12892-025-00289-0

Lundberg, S. M., Lee, S.-I. (2017). "A unified approach to interpreting model predictions," in Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). (Red Hook, NY, USA: Curran Associates Inc.), 4768-4777.

Nascimento, M., Peternelli, L. A., Cruz, C. D., Nascimento, A. C. C., Ferreira, R. D. P., Bhering, L. L., & Salgado, C. C. (2013). Artificial neural networks for adaptability and stability evaluation in alfalfa genotypes. Crop Breeding and Applied Biotechnology, 13, 152-156.

Novielli, P., Romano, D., Pavan, S., Losciale, P., Stellacci, A. M., Bellotti, R., & Tangaro, S. (2024). Explainable artificial intelligence for genotype-to-phenotype prediction in plant breeding: a case study with a dataset from an almond germplasm collection. Frontiers in Plant Science, 15. https://doi.org/10.3389/fpls.2024.1434229

Shukla, G. K. (1972). Some statistical aspects of partitioning genotype environmental components of variability. Heredity, 29(2), 237-245.

Tai, G. C. C. (1971). Genotypic stability analysis and its application to potato regional trials. Crop Science, 11(2), 184-190.

Teodoro, P. E., Barroso, L. M. A., Nascimento, M., Torres, F. E., Sagrilo, E., Santos, A. D., & Ribeiro, L. P. (2015). Redes neurais artificiais para identificar genótipos de feijão-caupi semiprostrado com alta adaptabilidade e estabilidade fenotípicas. Pesquisa Agropecuária Brasileira, 50(11), 1054-1060.

Tingxi Yu, Hao Zhang, Shoukun Chen, Shang Gao, Ze Liu, Jiankang Wang, Jose Crossa, Osval A Montesinos-López, Sarah Hearne, Huihui Li, EXGEP: a framework for predicting genotype-by-environment interactions using ensembles of explainable machine-learning models, Briefings in Bioinformatics, Volume 26, Issue 4, July 2025, bbaf414, https://doi.org/10.1093/bib/bbaf414

Wricke, G. (1962). Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. Zeitschrift für Pflanzenzüchtung, 47, 92-96.

Yan, W., Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Science, 40(3), 597-605.

Yao, Z., Yao, M., Wang, C., Li, K., Guo, J., Xiao, Y., ... & Liu, J. (2025). GEFormer: A genotype-environment interaction-based genomic prediction method that integrates the gating multilayer perceptron and linear attention mechanisms. *Molecular Plant*, *18*(3), 527-549.

Zhou, W., Yan, Z., & Zhang, L. (2024). A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction. Scientific Reports, 14, Article 5905. https://doi.org/10.1038/s41598-024-55243-x