Enhancing genomic prediction with Stacking Ensemble Learning in Arabica Coffee

Moyses Nascimento[1,2]*, Ana Carolina Campana Nascimento[1,2], Camila Ferreira Azevedo[1], Antonio Carlos Baiao de Oliveira[3], Eveline Teixeira Caixeta[3] and Diego Jarquin[2]*

[1]Laboratory of Intelligence Computational and Statistical Learning (LICAE), Department of Statistics, Federal University of Viçosa, Viçosa, Brazil, [2]Agronomy Department, University of Florida, Gainesville, FL, United States, [3]Embrapa Coffee, Brazilian Agricultural Research Corporation (Embrapa), Brasília, Brazil

Coffee Breeding programs have traditionally relied on observing plant characteristics over years, a slow and costly process. Genomic selection (GS) offers a DNA-based alternative for faster selection of superior cultivars. Stacking Ensemble Learning (SEL) combines multiple models for potentially even more accurate selection. This study explores SEL potential in coffee breeding, aiming to improve prediction accuracy for important traits [yield (YL), total number of the fruits (NF), leaf miner infestation (LM), and cercosporiosis incidence (Cer)] in Coffea Arabica. We analyzed data from 195 individuals genotyped for 21,211 single-nucleotide polymorphism (SNP) markers. To comprehensively assess model performance, we employed a cross-validation (CV) scheme. Genomic Best Linear Unbiased Prediction (GBLUP), multivariate adaptive regression splines (MARS), Quantile Random Forest (QRF), and Random Forest (RF) served as base learners. For the meta-learner within the SEL framework, various options were explored, including Ridge Regression, RF, GBLUP, and Single Average. The SEL method was able to predict the predictive ability (PA) of important traits in Coffea Arabica. SEL presented higher PA compared with those obtained for all base learner methods. The gains in PA in relation to GBLUP were 87.44% (the ratio between the PA obtained from best Stacking model and the GBLUP), 37.83%, 199.82%, and 14.59% for YL, NF, LM and Cer, respectively. Overall, SEL presents a promising approach for GS. By combining predictions from multiple models, SEL can potentially enhance the PA of GS for complex traits.

KEYWORDS

statistical and machine learning, prediction accuracy, plant breeding, ensemble methods, GBLUP

1 Introduction

Coffee is one of the most globally beverages presenting importance in terms of its potential health, socioeconomic, and economic effects (Porto et al., 2019). These effects drive breeding programs worldwide to develop high-yielding, adaptable cultivars delivering superior bean quality (Barbosa et al., 2019). However, traditional selection methods based on phenotypic observations of the plants or their family history (pedigree) are expensive and time consuming, especially for perennial crops as coffee.

An alternative approach denoted genomic selection (GS) has been used as a successful tool in genetic improvement (Meuwissen et al., 2001). GS helps increase genetic gain per generation by allowing for earlier selection through improved prediction of the potential of individual plants (Daetwyler et al., 2013; Nascimento et al., 2019; Voss-Fels et al., 2019). In the field of coffee breeding, GS has been utilized with the dual aim of accelerating genetic gain through early selection and improving prediction accuracy (Sousa et al., 2019; Alkimim et al., 2020; Sousa et al., 2021; Coelho de Sousa et al., 2022).

Among several prediction models, Genomic Best Linear Unbiased Prediction (GBLUP) is the most widely used approach for genomic prediction due to its advantages (Zhang et al., 2021). Compared to other parametric methods, GBLUP allows to accurately estimate narrow-sense heritability (Li et al., 2019) and presents higher computational efficiency (Hernandez et al., 2020). GBLUP modeling is also flexible. It can be modified to incorporate additional genetic information beyond the typical single-nucleotide polymorphism (SNP) markers. Specifically, this modeling allows to account for non-additive genetic effects, environmental factors, and even genotype-by-environment interactions, enriching the analysis and potentially improving prediction accuracy (Jarquín et al., 2014).

In the Artificial Intelligence Era, the interest in semi- and non-parametric methods for GS is increasing (Larkin et al., 2019; Coelho de Sousa et al., 2022; Seyum et al., 2022). These approaches, such as Artificial Neural Networks and Decision Trees, do not require prior assumptions about the relationships between inputs (SNP markers) and the output (phenotypic observations), allowing great flexibility to handle complex non-additive effects, such as dominance and epistasis (McKinney et al., 2006; Abdollahi-Arpanahi et al., 2020; Coelho de Sousa et al., 2022). In general, despite their potential, these approaches do not outperform the traditional parametric methods (e.g., GBLUP, Bayesian Alphabet - Gianola et al., 2009) used to predict the genetic merit of individuals (Liang et al., 2021).

Aiming to enhance predictive ability (PA), Ensemble Learning (EL) combines predictions from multiple models (base learners) into a single prediction (meta-learner) (Mendes-Moreira et al., 2012; Ganaie et al., 2022; Mienye and Sun, 2022). This approach leverages the strengths of diverse models to potentially generate more robust results compared to relying on a single learner (Liang et al., 2021; Kalule et al., 2023). In the context of GS, EL has found application through methods such as Random Forest (RF) and Bagging (Bag) (Xu et al., 2019; Abdollahi-Arpanahi et al., 2020; Sousa et al., 2021; Costa et al., 2022). These methods, categorized as Homogenous Learning (HL), utilize a single framework to produce

a single prediction value. Conversely, the Stacking Ensembles Learning (SEL) approach combines predictions from diverse methods, potentially outperforming HL (Mendes-Moreira et al., 2012). SEL has seen success in GS, improving PA in Chinese Simmental cattle, Dutch cattle, and pine (Liang et al., 2021), achieving higher accuracy than GBLUP for most evaluated traits.

Despite being interesting, EL approach arises some issues that needs to be considered. First, since the same individuals are used to fitting the model(s) in the EL approach, it is expected the existence of correlation between the predictions derived from the different methods. This well-known statistical problem is referred as multicollinearity (Montgomery et al., 2021) and causes high variability of the estimated effects. The second issue is related to which dataset should be used to fit the meta-learner. One option is to use directly the predicted values derived from the base learners. In this case, the simple mean or some regression model that accounts for multicollinearity problem (e.g., Ridge Regression) can be implemented to makeup a single prediction. An alternative option also could consider combining the predicted values with the genomic covariates (i.e., SNP markers|predicted values from the base learners) with the previous training data as new inputs. In this regard, in addition to the multicollinearity, the course of the dimensionality is another issue to consider mainly because these are covariables of different type.

Liang et al. (2021) used the predicted values derived from a multiple regression model as meta-learner. These authors obtained good results to improve PA compared to the conventional genomic prediction models on three different datasets. However, the use of an expanded training data augmented by SNP markers could be beneficial to further enhance the PA, and it emerges as an interesting approach. In this case, a model that addresses both multicollinearity and dimensionality problems should be used. One of the possible solutions can be considered to use a two-kernel GBLUP model as the meta-learner model. Another approach to evaluate is to consider only the predicted values provided by the best base learner models.

To date, no research has applied SEL to improve the prediction accuracy of important traits in coffee cultivars. This approach presents potential for coffee breeding, as it has been shown to outperform standard methods in other applications (Liang et al., 2021; Mohammed and Kora, 2023). By combining the strengths of multiple prediction models, SEL could lead to more reliable and accurate identification of valuable genetic traits in coffee plants, accelerating the development of superior coffee varieties.

In light of the mentioned points, the objective of this study was to use and evaluate the SEL to improve PA of important traits in Coffea Arabica. For that, the GBLUP, multivariate adaptive regression splines (MARS), Quantile Random Forest (QRF), and RF models were used as the base learner. Several approaches were considered as the meta-learner to construct the SEL framework. Specifically, the expanded- and non-expanded datasets were used for training. In addition, models that account for multicollinearity (Ridge Regression) and multicollinearity and dimensionality jointly (GBLUP) were also implemented.

2 Materials and methods

2.1 Phenotypic and genotypic data

The data were collected from the C. arabica breeding program, which is a joint partnership among the Agricultural Research Company of Minas Gerais (EPAMIG), the Federal University of Viçosa (UFV), and the Brazilian Agricultural Research Corporation (EMBRAPA). An experimental area is maintained at the Department of Phytopathology—UFV (20°44′25" S, 42°50′52" W). The database is composed of 13 progenies derived from crosses between three parents of the Catuaí cultivar and three parents of the Híbrido de Timor (HdT). Fifteen genotypes of the abovementioned progeny set (totaling 195 individuals) were genotyped for 21,211 SNP markers by Rapid Genomics, located in Gainesville Florida, USA. Information about the probes design and SNP identification are detailed in Sousa et al. (2017). The SNP markers set are widely distributed in the genome and in all coffee chromosomes, being useful for accurate studies on diversity and population structure, as well as selection and genomic association in C. arabica (Sousa et al., 2017, Sousa et al., 2019). The SNP quality control was carried out considering genotypic call rate and minor allele frequency equal to or greater than 90% and smaller than 5%, respectively. In this study a pre-selected set of 5,970 markers that did not reduce the PA of Arabica Coffee compared to the original set SNP markers in a previous study was used (Arcanjo et al., 2024).

The genotypes were planted on February 11, 2011, using a spacing of 3.0 m between rows and 0.7 m between plants following an augmented (check varieties) blocks experimental design. Nutritional management was carried out following the requirements of the crop. The phenotypic evaluations were performed in 2014, 2015, and 2016. A total of four traits were scored, two associated with the productivity, yield (YL—liters of fresh cherries harvested per plant) and total number of fruits (NF) —and two more associated with disease resistance—leaf miner infestation (LM) and cercosporiosis incidence (Cer) in Coffea Arabica. The incidence of cercosporiosis and leaf miner was evaluated using a score scale ranging from 1 to 5, in which 1 corresponded to genotypes without symptoms and 5 referred to highly susceptible genotypes. A comprehensive description of how the evaluations of each trait were performed can be found in Sousa et al. (2019).

2.2 Phenotypic data analysis

The phenotypic data for YL, NF, LM, and Cer were analyzed according to the following statistical model

$$\mathbf{y} = \mathbf{Xu} + \mathbf{Zg} + \mathbf{Wp} + \mathbf{Vr} + \mathbf{Tb} + \mathbf{Ri} + \mathbf{e}$$

where $\mathbf{y}$ represents the vector of observed phenotypes; $\mathbf{u}$ is the vector referring to the general mean in each evaluation year; $\mathbf{g}$ is the vector of genetic random effects corresponding to the progeny such that $\mathbf{g} \sim N(\mathbf{0}, \mathbf{I}\sigma_g^2)$; $\mathbf{p}$ is the random permanent environmental effect $\mathbf{p} \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$; $\mathbf{r}$ is the population random effect $\mathbf{r} \sim N(\mathbf{0}, \mathbf{I}\sigma_r^2)$; $\mathbf{b}$ is

the plot random effect $\mathbf{b} \sim N(\mathbf{0}, \mathbf{I}\sigma_b^2)$; $\mathbf{i}$ corresponds to the random effect of the interaction between progenies and the years $\mathbf{i} \sim N(\mathbf{0}, \mathbf{I}\sigma_i^2)$; and $\mathbf{e}$ is the experimental error assumed to be Independent and Identically Distributed (IID) outcomes from a normal density such that $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$. The genetic parameters, heritability and correlation, were also estimated for the evaluated traits. The individual heritability was estimated by $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_p^2 + \sigma_r^2 + \sigma_i^2 + \sigma_e^2}$. In addition, the adjusted phenotypes ($\mathbf{y}^*$, corrected BLUPs) for the year, plot, and year × progenies interaction effects were calculated and used for GS. The analyses were carried out using Selegen-REML/BLUP *software* (de Resende, 2016).

2.3 Individual genomic prediction

2.3.1 GBLUP

The parameterization of the Genomic prediction G-BLUP model can be defined as follows

$$\mathbf{y}^* = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where $\mathbf{y}^*$ is the vector of adjusted phenotypic observations as previously detailed; $\mathbf{b}$ is the vector of means; $\mathbf{X}$ is the incidence matrix corresponding to the fixed effects; $\mathbf{u}$ is the vector of individual additive genomic effects such that $\mathbf{u} \sim N(0, \mathbf{G}\sigma_g^2)$ where $\mathbf{G}$ is the kinship matrix describing genomic similarities between pairs of individuals, σ_g^2 is the additive genetic variance, $\mathbf{Z}$ is the incidence matrix that connect phenotypes with genotypes; $\mathbf{e}$ is the random error vector with $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$ where σ_e^2 is the residual variance. The additive genomic kinship matrix $\mathbf{G}$ was obtained as described by VanRaden (2008)

$$\mathbf{G} = \frac{\mathbf{W}^{\mathbf{T}}\mathbf{W}}{\sum_{i=1}^{n} 2p_i(1 - p_i)}$$

where, $\mathbf{W}$ is the centered (by columns) matrix of SNPs, which specifies the marker genotypes for each individual as 0, 1 or 2; p_i is the frequency of the second allele at the locus, that is,

$$W_{ij} = \begin{cases} 2 - 2p_j, & \text{if } M_{ij} = AA \\ 1 - 2p_j, & \text{if } M_{ij} = Aa \\ -2p_j, & \text{if } M_{ij} = aa \end{cases}$$

The *BGLR* function of the BGLR package (Pérez and de los Campos, 2014) in R software (R Core Team, 2022) was used to fitting GBLUP model.

2.3.2 Decision tree

The decision tree structure in this case is built using a regression tree algorithm. The objective is to create regions ($R_1, R_2, \ldots, R_M$) that minimize the difference between the predicted values and the adjusted observed values. This difference is measured by the Residual Sum of Squares (RSS). To achieve this, the algorithm performs a recursive binary splitting process. At each step, it considers all available features ($X_j -$ markers) and all possible split points (cutoff values) within each feature. The split that results in the lowest RSS for the resulting child nodes is chosen. This

process continues recursively until a stopping criterion is met, such as reaching a minimum number of data points in a region. Mathematically, the two disjoint regions can be defined by (Hastie et al., 2009)

$$R_1(j, s) = \{X|X_j < s\} \text{ and } R_2(j, s) = \{X|X_j \geq s\},$$

and the goal is to minimize:

$$\sum_{i\,:\,x_i \in R_1(j,s)} (y_i^* - \hat{y}_{R_1}^*)^2 + \sum_{i\,:\,x_i \in R_2(j,s)} (y_i^* - \hat{y}_{R_2}^*)^2$$

where $\hat{y}_{R_1}^*$ is the average of the adjusted phenotypic values of the training observations belonging to the region $R_1(j, s) = \{X|X_j < s\}$, $\hat{y}_{R_2}^*$ is the average of the adjusted phenotypic values of the training observations belonging to the region $R_2(j, s) = \{X|X_j \geq s\}$ and y_i^* is the true value of each individual.

2.3.3 Random Forest

To construct a RF is necessary to create several datasets by resampling (bootstrapping) from the training set. After that, the bootstrap samples are used to build multiple trees considering a subset of predictors (markers) randomly selected (Hastie et al., 2009). Usually, for a continuous response, the number of predictors used to find the best split at each node is a subset that is chosen by $m = \frac{v}{3}$, with v being the total number of predictors. Also, usually, the number of trees for the RF is set to 500. For the RF, the trees grow to their maximum size without pruning, and the prediction is done by averaging the trees. The function *randomForest* in randomForest R-package (Liaw and Wiener, 2002) was used to implement RF method.

2.3.4 Quantile Random Forest

For the construction of the QRF, as same as for RF, it is necessary to obtain T regression trees generated from bootstrap samples considering subsets of the markers under study (Hastie et al., 2009). Then, for the tth generated tree (T_t), the conditional distribution is obtained by weighting the observed values of the studied traits. Specifically, given an observation, $X = x$, it is defined for each terminal node (adjusted tree leaf), $F(x, T_{tf})$, the following weighting factor: $w_i(x, T_{tf}) = \frac{I_{\{x \in F(x, T_{tf})\}}}{\#\{m\,:\,X_m \in F(x, T_{tf})\}}$, with $\sum_{i=1}^n w_i(x, T_{tf}) = 1$, $I_{\{X_i \in F(x_i, T_{tf})\}}$ an indicator variable stating that the observed value ($X = x$) belongs to f^{th} leaf and $\#\{m\,:\,X_m \in F(x, T_{tf})\}$ represents the number of observations on the fth leaf.

The prediction of a tree T_t, according to Meinshausen (2006), for a new point, $X = x_{new}$, is given by the weighted average of the observations Y_i, that is, $\hat{\mu}(x_{new}) = \sum_{i=1}^n w_i(x, T_{tf})Y_i$. In this way, the prediction for a given observation, $X = x$, after the construction of T trees is given by $\hat{\mu}_{RF}(x) = \sum_{i=1}^n w_i(x)Y_i$ where $w_i(x) = \frac{1}{T}\sum_{t=1}^T w_i(x, T_{tf})$. Taking into consideration that the estimated cumulative distribution function is given by $\hat{F}(y|X = x) = \sum_{i=1}^n w_i(x)I_{\{Y_i \leq y\}}$, where $I_{\{Y_i \leq y\}}$ is an indicator function, the predicted value for the τth quantile is given by $Q_\tau(x) = \inf\{y : \hat{F}(y|X = x) \geq \tau\}$, for any τ, $0 < \tau < 1$.

The main difference between QRF and RF is that, for each node in each tree, the RF maintains only the average of the observations that fall into that node and discards any other information.

Conversely, the QRF maintains the value of all node observations (not just the average) and evaluates the conditional distribution based on this information (Meinshausen, 2006). This study evaluated nine Quantile Random Forest (QRF) for various quantile levels. The quantile parameter (τ) ranged from 0.1 to 0.9 in increments of 0.1. Therefore, the models were named QRF0.1, QRF0.2,…, QRF0.9, reflecting the specific quantile they aimed to predict. The function *quantregForest* in quantregForest R-package (Meinshausen, 2017) was used to implement the QRF methods.

2.3.5 Multivariate adaptive regression splines

MARS (Friedman, 1991) forms reflexive pairs of base functions (BF) for each input (marker) X_j, with nodes at each observed value x_{ij} of that input. The model building strategy is like a progressive linear regression, but instead of using the original inputs, it implements base functions from the set $C = \{(X_j - t)_+,\ (t - X_j)_+\}_{t \in \{x_{1j},\ x_{2j},\ ...,\ x_{Nj}\}\ j=1,2,...,p}$ and/or its products. The MARS model, which is a linear combination of the BF and/or their interactions, is given by (Hastie et al., 2009):

$$f(X) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X)$$

where β_0 is the regression constant, β_m with $m = 1, 2,…, M$, are the regression coefficients, and $h_m(X)$ is a function in C, or a product of two or more functions.

The estimation process of the parameters β_0 and β_m is based on the minimization of the residual sum of squares. First, the forward phase starts on the training data, building the model initially with only the constant function $h_0(X) = 1$, and all functions in the C set are candidate functions. At each subsequent step, the base pair that produces the maximum reduction in training error is added. Considering a model with basic M functions, the next pair to be added to the model is

$$\hat{\beta}_{M+1}h_l(X)(X_j - t)_+ + \hat{\beta}_{M+2}h_l(X)(t - X_j)_+,\ h_l \in M$$

where $\hat{\beta}_{M+1}$ and $\hat{\beta}_{M+2}$ are coefficients estimated by the least square method (Hastie et al., 2009), together with all other M + 1 coefficients in the model. This process of adding BF continues until the model reaches a predetermined maximum number, often leading to a purposefully overparametrized model (Zhang and Goh, 2016). The backward phase improves the model by removing the least significant terms until finding the best sub model. The model subsets are compared using the generalized cross-validation (GCV) method. The GCV is evaluated with the root-mean-square residual error divided by a penalty that depends on the complexity of the model (Zhang and Goh, 2016) and it is calculated as

$$GCV(\lambda) = \frac{\frac{1}{N}\sum_{i=1}^N [y_i - \hat{f}_\lambda(x_i)]^2}{\left[1 - \frac{C(M)}{N}\right]^2}$$

(Hastie et al., 2009) where M is the effective number of model parameters, C(M) is a cost function for each basis function included in the developed submodel, which by default adopts the value of 3, N is the number of datasets used in CV and $\hat{f}_\lambda(x_i)$ denotes the predicted MARS values. This study employed three Adaptive

Regression Spline (MARS) models with varying degrees of interaction (1, 2, and 3). The MARS 1 model represents an additive model, meaning it captures only the linear effects of markers. In contrast, MARS 2 and MARS 3 allow for the inclusion of second and third-order interactions, respectively, enabling them to capture more complex non-additive relationships between markers. The function *earth* in *earth* R-package (Milborrow, 2017) was used to implement MARS models.

2.4 Stacking Ensemble Learning for genomic prediction

This study explores the SEL approach for improving the accuracy of genomic prediction models. SEL leverages predictions from multiple individual models (base learners, Level 0) and combines them using a meta-learner model (Level 1) to generate a final, potentially more accurate prediction. The base learners used in this study were GBLUP, different degrees of MARS (1, 2, 3), multiple QRFs (0.1 to 0.9), and RF. Their predictions, referred to as Genomic Estimated Breeding Values from Base Learners (GEBV-BL), formed the core metadata for the meta-learner. In practice, it is necessary to prepare a dataset with both the observable characteristics (phenotype) and the genetic information (SNP markers) of individuals. Then, diverse machine learning models (base learners) are trained on these data to make initial predictions.

These predictions from the base learners become the new features for a final model, the meta-learner. Finally, the meta-learner is trained using these base learner predictions as input and the original phenotype data as the target variable. In our work, four different combinations of metadata were explored: (i) GEBV-BL, only predictions from the base learners (standard approach); (ii) GEBV-BL+SNP, predictions combined with the original Single Nucleotide Polymorphism (SNP) markers (larger input dataset); (iii) GEBV-BL-Best; and (iv) GEBV-BL-Best + SNP, Similar to the previous cases, but only predictions from high-performing base learners (those exceeding the average predictive accuracy) were included. For GEBV-BL and GEBV-BL-Best datasets, six meta-learner methods were evaluated: Simple Mean (SSM); Weighted Regression (SWR); Regression (SR); Ridge Regression (SRR); Random Forest (SRF). For GEBV-BL + SNP and GEBV-BL-Best + SNP datasets, which included SNP markers, a two-kernel GBLUP model (S2KGBLUP) was additionally employed as the meta-learner. The SEL scheme for genomic prediction is illustrated in the Figure 1.

2.5 Cross-validation

The PA of the models used as base-learners and the entire SEL process considered a CV scheme that was implemented as follows. First, the complete dataset under study was randomly divided into
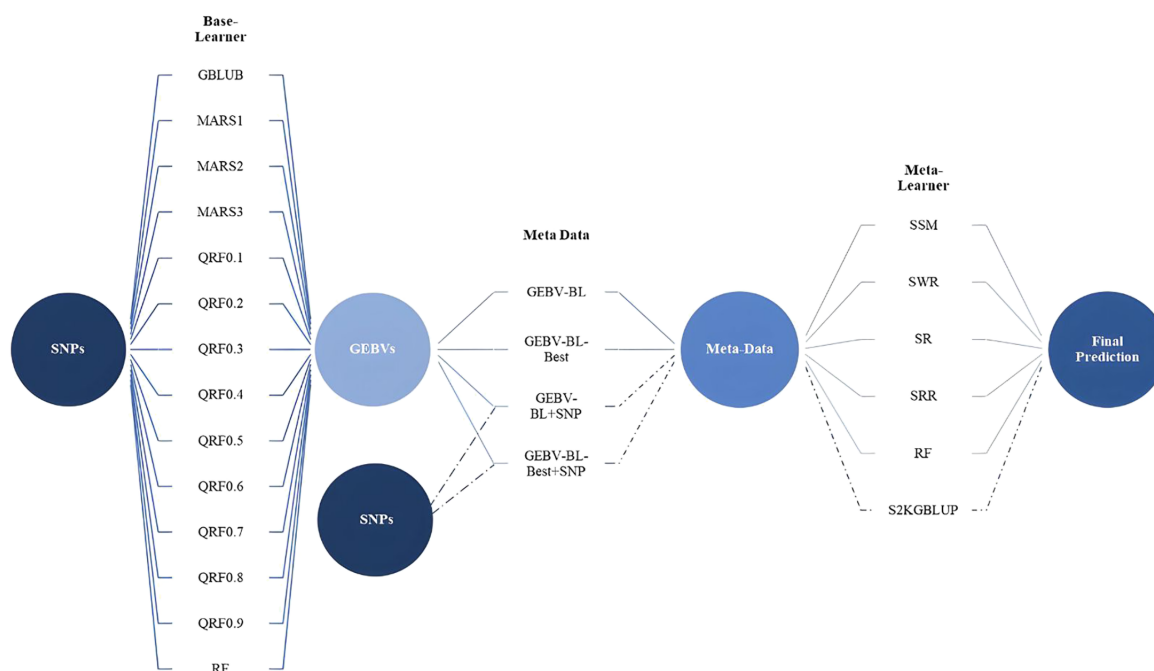

FIGURE 1
The stacking ensemble learning framework for genomic prediction from original data to the base learners, creating metadata for the meta-learner. Base-Learner (Level 0) is composed of the GBLUP, MARS (1°, 2°, and 3°), QRF considering nine quantiles (from 0.1 to 0.9, in steps of 0.1) and RF model. Four different meta-data were obtained: (i) GEBV-BL, only predictions from the base learners (standard approach); (ii) GEBV-BL+SNP, predictions combined with the original Single Nucleotide Polymorphism (SNP) markers (larger input dataset); (iii) GEBV-BL-Best; and (iv) GEBV-BL-Best + SNP, similar to the previous cases, but only predictions from high-performing base learners (those exceeding the average predictive accuracy). Meta-Learners: for GEBV-BL and GEBV-BL-Best metadata, six meta-learner methods were evaluated. Simple Mean (SSM), Weighted Regression (SWR), Regression (SR), Ridge Regression (SRR), Random Forest (SRF), and for GEBV-BL + SNP and GEBV-BL-Best + SNP datasets, which included SNP markers, a two-kernel GBLUP model (S2KGBLUP).

two sets (training and testing). The training set was composed by 70% of the individuals while the remaining 30% was assigned to the testing or validation set. The training set was used to calibrate the base-learners and the SEL for predicting the GEBVs of the individuals in the training set. This procedure was repeated 10 times. Then, for each approach, the average PA across replicated was computed. The PA was computed as the Pearson correlation between predicted GEBV and the adjusted phenotype values. The standard error (SE) was also computed. In addition, the mean square error (MSE) between the observed and predicted values was calculated. Finally, the agreement coefficient was used to compute the percentage of individuals with a performance above the 90th percentile in fields given the top 10% of the GEBVs obtained with the different genomic prediction approaches.

3 Results

3.1 Phenotypic data analysis

The across environments mean ($\bar{X}$) and standard deviation (SD) of the evaluated traits are summarized in Table 1.

The estimates of the heritability (proportion of phenotypic variability explained by the genetic component) for YL (0.30), NF (0.49), LM (0.30), and Cer (0.38) were moderate. The Spearman's correlation (lower triangle) between the adjusted phenotypic values of each pair of traits were positive and presented low to moderate values varying from 0.02 to 0.52. The higher and the lower correlation values were observed between YL and NF (0.52) and between NF and Cer (0.02, not statistically significant), respectively (Figure 2). The correlation between YL and LM, Cer and NF and LM, Cer were not statistically significant (Supplementary Figure S1).

3.2 Comparison between the base learners

Overall, none of the evaluated base learner methods outperformed the predictive performance of the others for all the evaluated traits. The estimated predictive abilities (PA) and corresponding standard deviations for the four traits (YL, NF, LM, and Cer) ranged from −0.01 (0.01) to 0.24 (0.01) and are presented in Figure 3. Specifically, for YL, NF, LM, and Cer, the

highest PA values were 0.15 (0.01), 0.24 (0.01), 0.15 (0.01) and 0.24 (0.02), and these were obtained with MARS2 and QRF0.3, QRF0.7 and GBLUP methods, respectively (Figure 2).

GBLUP presented lower values of EQM (Supplementary Table S1). Specifically, the MSE were equal to 16.05 (1.01), 2,329.00 (368.94), 0.21 (0.02), and 0.57 (0.07) for YL, NF, LM, and Cer, respectively.

The extreme QRF models QRF0.1, and QRF0.9 returned the highest MSE values across all the evaluated traits (Supplementary Table S1). In general, the MSE decreased as the fitted quantile model was approaching to the median model (QRF0.5).

3.3 Comparison between the Stacking Ensemble Learning approaches and GBLUP

The estimates of the PA obtained with the SEL models and the traditional genomic prediction method GBLUP model are shown in Figure 3. The results of the GBLUP model were used as benchmark since as it was mentioned it is the most convenient and used implementation in genomic prediction.

The estimated PA ranged from 0.05 (0.01) to 0.32 (0.02) (Figure 4). For YL, NF, LM, and Cer, the highest PA values were 0.20 (0.01), 0.32 (0.01), 0.19 (0.01), and 0.27 (0.01), respectively. These results were obtained by implementing SMBest method, which corresponds to the simple mean considering GEBV-BL-Best metadata. The "best" fitted model was SMBest, and it outperformed the PA of the GBLUP model by 87.44% (the ratio between the PA obtained from SMBest model fit and the GBLUP), 37.83%, 199.82%, and 14.59% for YL, NF, LM, and Cer, respectively (Figure 3).

Regarding the different data sets used as input in the SEL approach, combining the predicted values obtained from base learners (GEBV-BL) with training data used to fitting the models, did not improve PA of these methods (Figure 3). Additionally, the results considering only the predicted values provided from those base learners with PA higher than mean of all base learner (GEBV-BL-Best) as input in the Level 1, returned the highest results (Figure 3).

For the four traits, the GBLUP model presented the lowest MSE values (Supplementary Table S2). As expected, since the Ridge Regression model (SRR) depends on a regularization parameter it presented a significant higher MSE (Supplementary Table S2). For this model, the MSE values were equal to 69.28 (9.18), 8390.68 (518.49), 6.65 (0.80), and 3.45 (0.36) for YL, NF, LM, and Cer, respectively.

The Spearman's correlation between the GEBVs obtained with the different prediction models, including the baseline GBLUP model and all the SEL models, presented positive values and these vary from low 0.25 to high 0.97 across the evaluated traits (Figures 4–7, lower triangular matrix). Low values of the Spearman's correlation were observed between the GBLUP and the other SEL methods and these were 0.25, 0.33, 0.28, and 0.30 for YL, NF, LM, and Cer, respectively (Figures 4–7, lower triangle). On the other hand, the highest correlation value (0.97) was observed between the GEBVs obtained by the Stacking Regression (SR) and the Stacking Ridge Regression (SRR) for LM (Figure 6).

TABLE 1 Across environments phenotypic mean ($\bar{X}$) and standard deviation (SD) for yield (YL), total number of fruits (NF), leaf miner infestation (LM), and cercosporiosis incidence (Cer) of a coffea arabica L. population composed of 195 individuals observed in years 2014, 2015, and 2016 in Viçosa, Brazil.

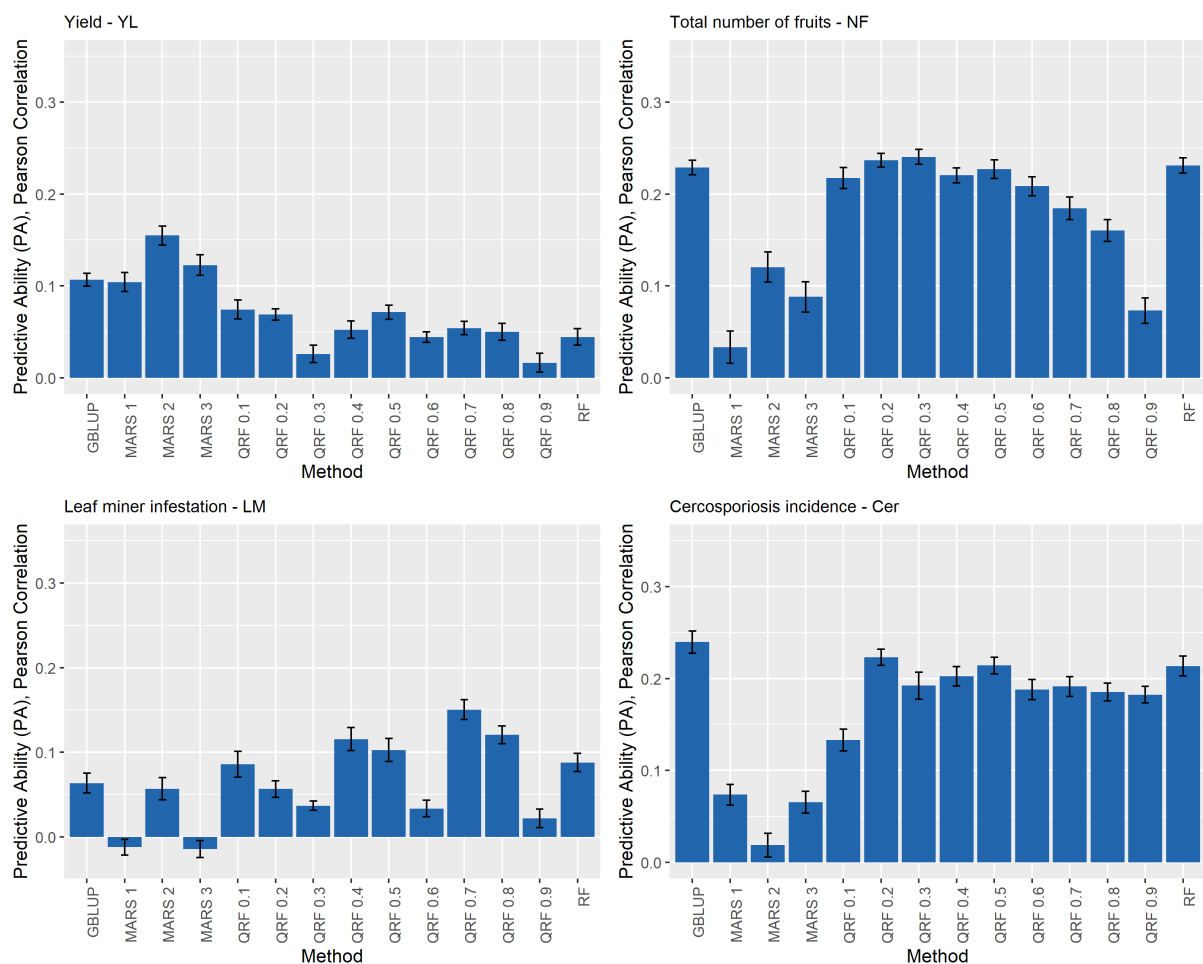| Trait	$\bar{X}$	SD
YL	5.16	3.84
NF	2.32	0.55
LM	2.05	0.69
Cer	2.48	0.70

FIGURE 2
Predictive ability (PA) for yield (YL), total number of fruits (NF), leaf miner infestation (LM), and cercosporiosis incidence (Cer) measured in an Arabica coffee population composed of 195 individuals using a holdout validation scheme repeated 10 times. The fitted models used as base learners are: Genomic Best Linear Unbiased Predictor (GBLUP); Multivariate Adaptive Regression Splines with degrees equal to 1, 2, and 3 (MARS 1, MARS 2 and MARS 3); Quantile Random Forest evaluated at nine quantiles [(τ): 0.1 to 0.9, every 0.1] − (QRF 0.1, …, QRF 0.9), and Random Forest (RF).

For each prediction method, the predicted values were ordered based on rankings then the percentage of common individuals in the top 10% between pairs of methods was computed. Overall, the GBLUP presented lower agreement with the SEL evaluated approaches (Figures 4–7, upper triangle). For instance, the agreement coefficient between the GBLUP and the SRB, SR, SR, and S2KGBLUP methods presented values varying from 0.31 to 0.48 for all evaluated traits (Figures 4–7, upper triangle).

Regarding the different data sets used as input in the SEL approach, the highest Spearman's correlations and agreements were observed between those methods that used the same kind of metadata as input in the fitting. Overall, considering these two measures, the methods were grouped into three groups (Supplementary Figures S6-S13). In general, the GBLUP was allocated into a single group. The only exception was for cercorporiosis (Cer) considering the agreement measure. Is this case, the GBLUP was allocated together with those methods that's considers only the predicted values provided from those base

learners with PA higher than the mean of all base learner (GEBV-BL-Best) as input in the SEL approaches (Supplementary Figure S12).

4 Discussion

In this study, we used the SEL approach to improve PA of four important traits in Coffea Arabica. Two of these traits are associated with the productivity (YL and NF) and the remaining two with disease resistance (LM and Cer). The population under study is comprise of 195 genotypes of Coffea Arabica genotyped for 5,970 SNP markers. We compared the PA of different approaches used in the Level 1 of the SEL to the results obtained with the base learners [GBLUP, MARS considering degrees equal to 1, 2, and 3, QRF considering nine quantiles (from 0.1 to 0.9, every 0.1) and RF]. Since the GBLUP is the most implemented prediction model (Zhang et al., 2021), their results were used as benchmark. The

PA of the different approaches was assessed using a CV scheme repeated 10 times. The Spearman's correlation and the agreement (based on the top 10%) coefficients between the GEBV values of the different models were also estimated. The genetic parameters were also estimated for the evaluated traits (YL, NF, LM, and Cer).

The heritability estimates for YL (0.55), NF (0.49), LM (0.30), and Cer (0.38) were consistent with those reported in the literature for this specie and same traits. Specifically, the heritability estimates varied between (0.1–0.74) [Alkimim et al. (2020) and Alemayehu, 2019], (0.30–0.55) [Gokavi et al., 2023 and Weldemichael et al., 2017], (0.30–0.51) [Chrigui et al., 2020 and Ferrão et al., 2023], and (0.09–0.61) Alkimim et al., 2021 and Ferrão et al., 2023] for the YL, NF, LM, and Cer, respectively. Although the Pearson correlation between YL and the disease resistance traits were not statistically significant, a significant and positive genetic correlation (0.52) was obtained between YL and NF.

The machine learning methods as base learners have been already used in genomic prediction (Long et al., 2011; Lenz et al.,

2019; Montesinos-López et al., 2019; Coelho de Sousa et al., 2022; Costa et al., 2022). However, generally these methods do not outperform significantly the traditional genomic prediction approach based on parametric models such as GBLUP and Bayesian Alphabet (Liang et al., 2021).

Liang et al. (2021) used the SEL for improving PA in three real datasets on average by 7.70%, compared to GBLUP. The SEL uses predicted values from different machine learning implementations to obtain a single prediction value. These authors integrated/ combined the results of three machine learning implementations (Support Vector Machine, Kernel Ridge Regression and Elastic Net) to compute the GEBVs.

In contrast to Liang et al. (2021), in our study, the GBLUP approach was used as one of the base learner methods for the SEL too. The GBLUP was considered since it is widely used for genome prediction (Zhang et al., 2021) due to its reduced computational demand and simplicity (Hernandez et al., 2020) compared to the other parametric methods (e.g., Bayesian Alphabet, Gianola et al.,
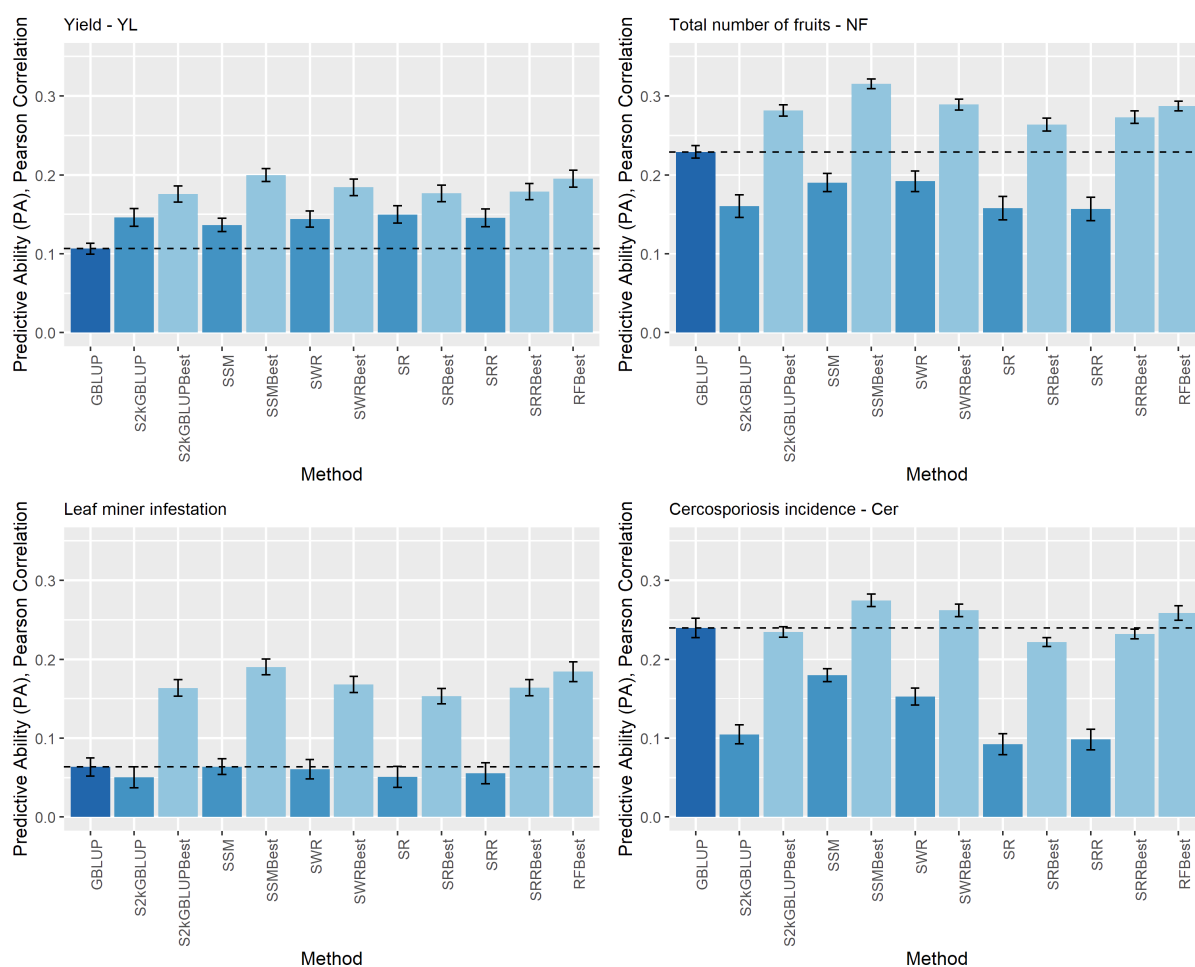

FIGURE 3
Predictive ability (PA) for yield (YL), total number of fruits (NF), leaf miner infestation (LM), and cercosporiosis incidence (Cer) measured in an Arabica coffee population composed of 195 individuals using a holdout validation scheme repeated 10 times. The fitted models used as base learners are: Stacking Simple Mean (SSM), Stacking Weighed Regression (SWR), Stacking Regression (SR), Stacking Ridge Regression (SRR), and the Stacking two-kernel GBLUP model (S2KGBLUP). The models named as best (SSMBest, SWRBest, SRBest, SRRBest, S2KGBLUP, and RFBest) used in the fitting only the results provided by those methods that presented predictive ability higher the mean in the Level 0.

2009). The MARS that allows automatically selecting and modeling nonlinear relationships and interaction effects of the predictor variables was also considered as base-learner method (Costa et al., 2022). In addition, the RF (James et al., 2023) and the QRF (Meinshausen, 2006) were also set as base learner methods. Specifically, the RF is a machine learning approach used to increase the predictive power and reduce the variance of the predicted values by averaging uncorrelated quantities (James et al., 2023). The QRF combines interesting characteristics from RF and Quantile Regression (QR) approaches. QR (Koenker and Bassett, 1978) allows fitting all the portions of the probability distribution of the trait, enabling a more complete picture of the conditional distribution than a single estimate of the center (Briollais and Durrieu, 2014; Nascimento et al., 2019).

Overall, for each evaluated trait (YL, NF, LM, and Cer), a different model presented the highest PA. These results show that there is not a single approach that outperforms the others in the evaluated data sets. Also, it could be case of a model performing better than the others in a given dataset but another model could perform better in a similar dataset (James et al., 2023). For example,

the MARS models with 2° and 3° (model with interactions) presented higher PA for YL. These results are in line with those obtained by Coelho de Sousa et al. (2022). Using artificial neural networks to predict the genetic merit of genotypes of *Coffea canephora* these authors showed a larger dominance markers effect for YL when compared to the GBLUP additive dominant model based on additive marker effects.

Another interesting result was obtained modeling the QRF where the distribution of the adjusted phenotypic values for YL and NF. These presented a higher concentration in the first quantiles (0.1–0.3) (Supplementary Figure S2) and positive skewness (Supplementary Figure S3). For YL, the best models were the QRF0.1 and QRF0.2, and QRF0.3 for NF (Supplementary Figure S2). The distribution of the adjusted LM values presented tree major portions (Supplementary Figure S4). The QRF modeling was able to distinguish these three different groups (Figure 2). Finally, since the distribution of adjusted Cer phenotypic values did no present a specific pattern to highlight (Supplementary Figure S5) all of the QRF models present similar PA (Figure 2). A similar trend was shown by Nascimento et al.
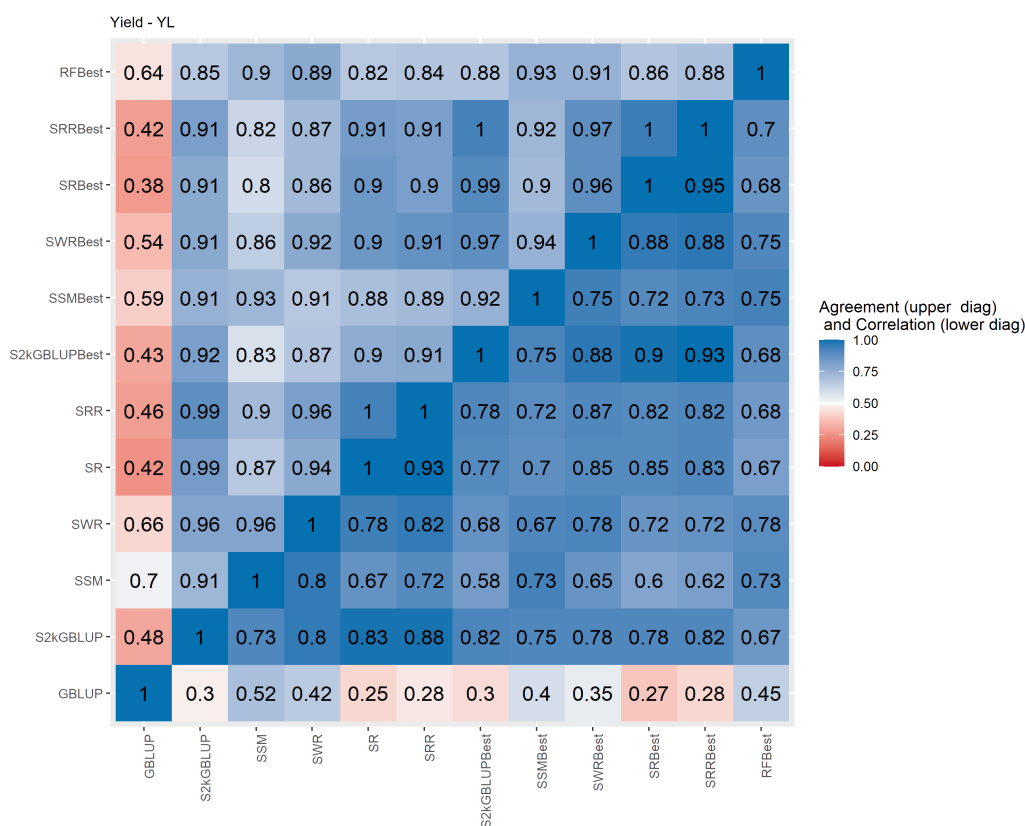

FIGURE 4
Spearman's correlation between the genomic estimated breeding values [GEBV] (lower diagonal matrix) and the concordance coefficient between the top 10% of the selected individuals (upper triangular matrix) considering all the different fitted models including the GBLUP model and the meta-learners for yield (YL). The fitted models used as base learners are: Stacking Simple Mean (SSM), Stacking Weighed Regression (SWR), Stacking Regression (SR), Stacking Ridge Regression (SRR), and the Stacking two-kernel GBLUP model (S2KGBLUP). The models named as best (SSMBest, SWRBest, SRBest, SRRBest, S2KGBLUP, and RFBest) used in the fitting only the results provided by those methods that presented predictive ability higher the mean in the Level 0.

(2017). These authors found that the Quantile Regression approach outperform the traditional genomic prediction methods of not normal distributed traits.

An interesting approach to address the non-normality assumption is using multiple models to conduct the predictions, and then combine the predicted values to makeup a single prediction through the SEL approach. In general, the SEL outperforms the methods based on base learners only (Liang et al., 2021; Kandel et al, 2021; Kalule et al., 2023). In our study, the SEL approach outperformed all base learner methodologies (Figures 2, 3). However, it is important to emphasize that these results were observed by those SEL models that used only the predicted values provided from the base learners with PA higher than mean of all the base learners. Specifically, the Stacking Mean Best (SMB) presented the highest PA for all of the evaluated traits. The average of the predictions from several fitted models has been successfully implemented with Bagging and RF approaches (Breiman, 1996 and Breiman, 2001). The SEL approach allow to use several models to combine the predicted values, for example, XGBoost (Ghasemieh et al., 2023), Penalized methods (Kalule et al.,

2023), Linear Regression (Liang et al., 2021). Similar to the single model approach, the performance of the different SEL implementation can vary from one data set to another with no one of these outperforming the others in all data sets. Thus, as it was recommended by these authors it is important to evaluate several models as meta-learners as well.

Regarding the MSE, as expected, the penalized models, showed larger values compared to the other evaluated methods. By design these methods induce bias aiming to reduce the variance of the estimations (Montgomery et al., 2021; Chan et al., 2022). However, these cannot guarantee the increasing of the PA compared to other methods. The SMB, which resulted to return the best results in terms of PA, also presented large values for the MSE. This can be a consequence that SMB-SEL uses predicted values derived from base learners that return large MSE values (Supplementary Table S1).

Overall, the SEL models presented moderate to high Spearman's correlation between them (Figures 4–6). On the other hand, these were low to moderate between SEL approaches and the GBLUP model. Additionally, among the 10% of genotypes with the highest GEBVs for YL, NF, LM, and Cer, the agreement coefficient between
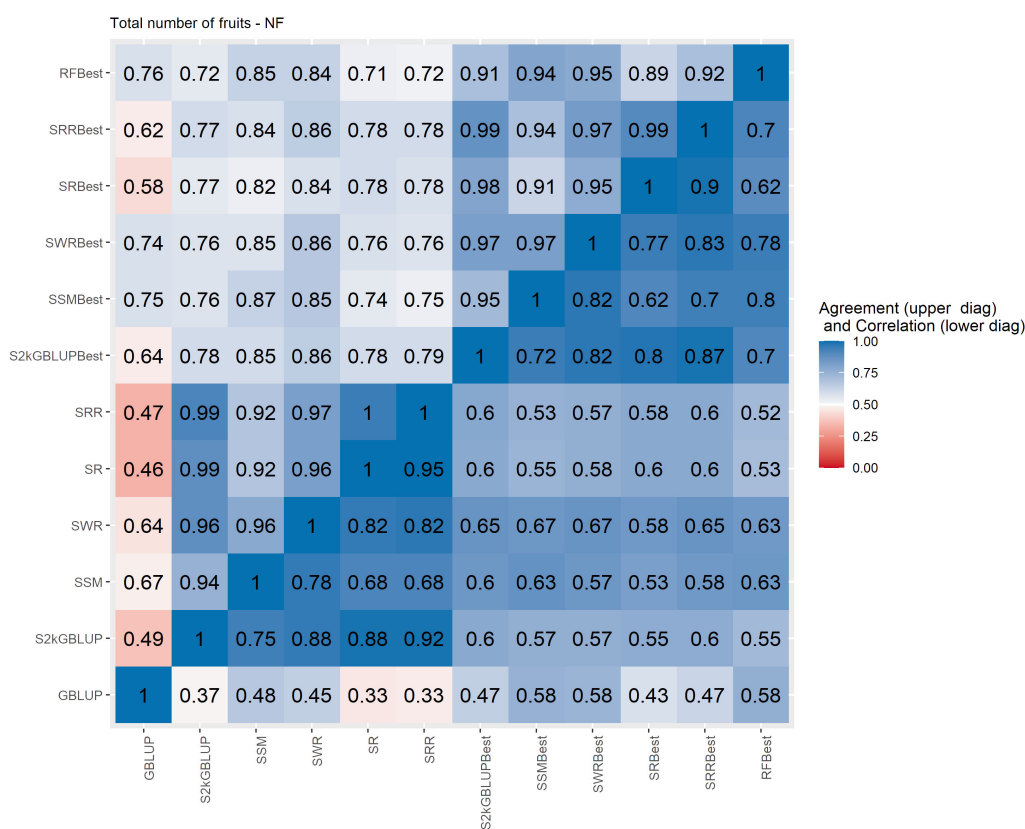

FIGURE 5
Spearman's correlation between the genomic estimated breeding values [GEBV] (lower diagonal matrix) and the concordance coefficient between the top 10% of the selected individuals (upper triangular matrix) considering all the different fitted models including the GBLUP model and the meta-learners for total number of fruits (NF). The fitted models used as base learners are Stacking Simple Mean (SSM), Stacking Weighed Regression (SWR), Stacking Regression (SR), Stacking Ridge Regression (SRR), and the Stacking two-kernel GBLUP model (S2KGBLUP). The models named as best (SSMBest, SWRBest, SRBest, SRRBest, S2KGBLUP, and RFBest) used in the fitting only the results provided by those methods that presented predictive ability higher the mean in the Level 0.

the SEL and GBLUP models showed values varying from moderate to high, suggesting differences in the obtained classifications with these. In general, the cluster analysis of these results showed that the methods can be grouped into three distinct groups (Supplementary Figures S6-S13) with the GBLUP forming a group by itself.

Altogether, these results show that the use of SEL to predict the individual genetic merit of four important traits in Arabica Coffee is worth to investigate. The SEL approach showed higher estimates of PA compared with all evaluated base learning methods, in special to the traditional GBLUP method. In practice, SEL's ability to combine methods with diverse characteristics facilitates a more comprehensive exploration of the relationships between variables leading to more accurate selection of breeding. This approach considers a wider range of factors and reduces the reliance on any single model's limitations. However, evaluating phenotypes across multiple environments can pose challenges for SEL. Unlike GBLUP, which presents higher computational efficiency (Hernandez et al., 2020), many base learners in SEL methods are based on machine learning requiring significant computation time in certain scenarios. Studies have explored the use of single machine learning methods for multi-environment trials (METs). For example, Barreto et al. (2024) applied machine learning to predict hybrid performance in METs and

achieved similar PA compared to GBLUP with non-additive effects. As highlighted by Montesinos López et al. (2022) in their study using RF for METs, training any machine learning model can be computationally demanding, especially for datasets where the training data sets are very large. The hyperparameter tuning for individual base learners within a SEL framework is a well-established approach to enhance model performance. However, it is important to acknowledge that SEL ensembles can achieve strong results even with default base learner parameters (Friedel et al., 2023). This aligns perfectly with the core principle of ensemble learning, that is, leveraging predictions from multiple models can outperform any single model. In our study, the high dimensionality of the data presented significant computational challenges for hyperparameter tuning. Additionally, the observed superiority of the SEL approach compared to traditional methods suggested that tuning might not be as critical for achieving good results.

5 Conclusion

The SEL method was able to predict the PA of important traits (YL, NF, leaf miner infesting, cercosporiosis resistance) in Coffea
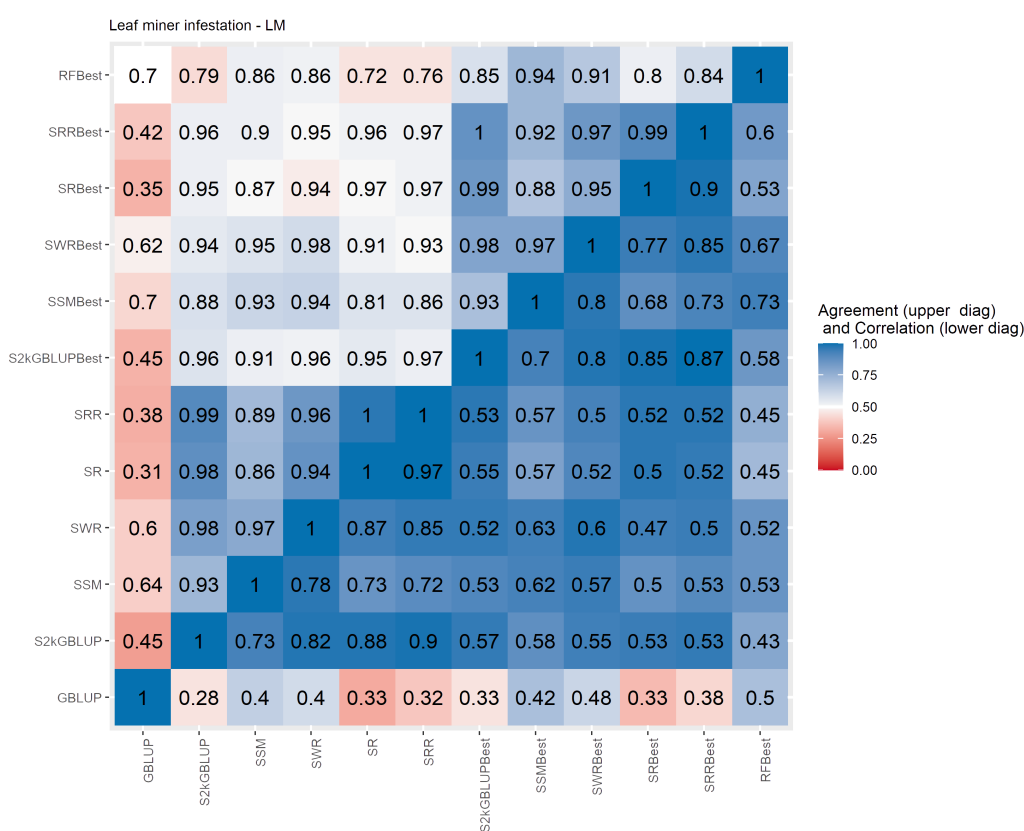

FIGURE 6
Spearman's correlation between the genomic estimated breeding values [GEBV] (lower diagonal matrix) and the concordance coefficient between the top 10% of the selected individual's upper triangular matrix) considering all the different fitted models including the GBLUP model and the meta-learners for leaf minor infestation (LM). The fitted models used as base learners are: Stacking Simple Mean (SSM), Stacking Weighed Regression (SWR), Stacking Regression (SR), Stacking Ridge Regression (SRR), and the Stacking two-kernel GBLUP model (S2KGBLUP). The models named as best (SSMBest, SWRBest, SRBest, SRRBest, S2KGBLUP, and RFBest) used in the fitting only the results provided by those methods that presented predictive ability higher the mean in the Level 0.
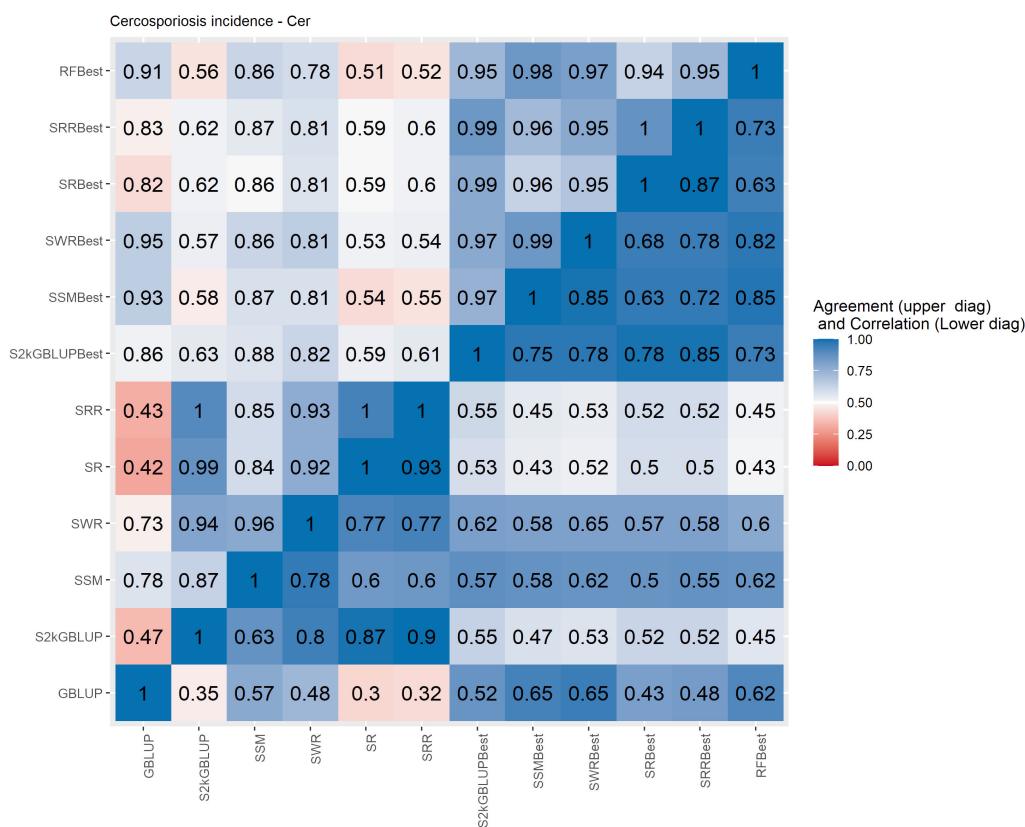
FIGURE 7
Spearman's correlation between the genomic estimated breeding values [GEBV] (lower diagonal matrix) and the concordance coefficient between the top 10% of the selected individuals (upper triangular matrix) considering all the different fitted models including the GBLUP model and the meta-learners for Cercosporiosis incidence (Cer). The fitted models used as base learners are: Stacking Simple Mean (SSM), Stacking Weighed Regression (SWR), Stacking Regression (SR), Stacking Ridge Regression (SRR), and the Stacking two-kernel GBLUP model (S2KGBLUP). The models named as best (SSMBest, SWRBest, SRBest, SRRBest, S2KGBLUP, and RFBest) used in the fitting only the results provided by those methods that presented predictive ability higher the mean in the Level 0.

Arabica. In addition, SEL presented higher PA compared with those obtained for all base learner methods (GBLUP, MARS considering degrees equal to 1, 2, and 3, QRF considering nine quantiles, from 0.1 to 0.9, every 0.1 and RF).

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

MN: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. AN: Writing – review & editing, Software,

Methodology, Investigation. CA: Writing – review & editing, Methodology. AO: Writing – review & editing, Data curation. EC: Writing – review & editing, Investigation, Data curation. DJ: Writing – review & editing, Methodology, Investigation.

respectively, from Brazilian Council for Scientific and Technological Development (CNPq).

Conflict of interest

Authors AB and EC were employed by the company Embrapa Coffee, Brazilian Agricultural Research Corporation (Embrapa).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1373318/full#supplementary-material

References

Abdollahi-Arpanahi, R., Gianola, D., and Peñagaricano, F. (2020). Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genet. Selection Evol.* 52. doi: 10.1186/s12711-020-00531-z

Alemayehu, D. (2019). Estimation of genetic component and heritability for quantitative traits in amaro coffee (Coffea Arabica L.) landrace at Awada, Southern Ethiopia. *Int. J. Res. Stud. Science Eng. Technology.* 6, 1–9.

Alkimim, E. R., Caixeta, E. T., Sousa, T. V., Gois, I. B., Lopes da Silva, F., Sakiyama, N. S., et al. (2021). Designing the best breeding strategy for Coffea Canephora: Genetic Evaluation of pure and hybrid individuals aiming to select for productivity and disease resistance traits. *PLoS One* 16. doi: 10.1371/journal.pone.0260997

Alkimim, E. R., Caixeta, E. T., Sousa, T. V., Resende, M. D., da Silva, F. L., Sakiyama, N. S., et al. (2020). Selective efficiency of genome-wide selection in Coffea canephora breeding. *Tree Genet. Genomes* 16. doi: 10.1007/s11295-020-01433-3

Arcanjo, E. S., Nascimento, A. C. C., Nascimento, M., Azevedo, C. F., Caixeta, E. T., Oliveira, A. C. C., et al. (2024). Low-density marker panels for genomic prediction in Coffea arabica L. Acta Scientiarum. *Agronomy.*

Barbosa, I. D. P., da Costa, W. G., Nascimento, M., Cruz, C. D., and de Oliveira, A. C. (2019). Recommendation of Coffea arabica genotypes by factor analysis. *Euphytica* 215. doi: 10.1007/s10681-019-2499-x

Barreto, C. A. V., das Graças Dias, K. O., de Sousa, I. C., Azevedo, C. F., Nascimento, A. C. C., Guimarães, L. J. M., et al. (2024). Genomic prediction in multi-environment trials in maize using statistical and machine learning methods. *Sci. Rep.* 14, 1062. doi: 10.1038/s41598-024-51792-3

Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 24, 123–140. doi: 10.1007/BF00058655

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Briollais, L., and Durrieu, G. (2014). Application of quantile regression to recent genetic and -omic studies. *Hum. Genet.* 133, 951–966. doi: 10.1007/s00439-014-1440-6

Chan, J. Y.-L., Leow, S. M., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z.-W., et al. (2022). Mitigating the multicollinearity problem and its Machine Learning Approach: A Review. *Mathematics* 10, 1283. doi: 10.3390/math10081283

Chrigui, N., Sari, D., Sari, H., Eker, T., Cengiz, M. F., Ikten, C., et al. (2020). Introgression of resistance to Leafminer (Liriomyza Cicerina Rondani) from Cicer reticulatum Ladiz. to C. Arietinum L. and relationships between potential biochemical selection criteria. *Agronomy* 11, 57. doi: 10.3390/agronomy11010057

Coelho de Sousa, I., Nascimento, M., de Castro Sant'anna, I., Teixeira Caixeta, E., Ferreira Azevedo, C., Damião Cruz, C., et al. (2022). Marker effects and heritability estimates using additive-dominance genomic architectures via artificial neural networks in Coffea canephora. *PLoS One* 17, e0262055. doi: 10.1371/journal.pone.0262055

Costa, W. G., Celeri, M. O., Barbosa, I. P., Silva, G. N., Azevedo, C. F., Oliveira, A. B., et al. (2022). Genomic prediction through machine learning and neural networks for traits with epistasis. *Comput. Struct. Biotechnol. J.* 20. doi: 10.1016/j.csbj.2022.09.029

Daetwyler, H. D., Calus, M. P., Pong-Wong, R., de los Campos, G., and Hickey, J. M. (2013). Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. *Genetics* 193, 347–365. doi: 10.1534/genetics.112.147983

de Resende, M. D. V. (2016). Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breed. Appl. Biotechnol.* 16, 330–339. doi: 10.1590/1984-70332016v16n4a49

Ferrão, M. A., da Fonseca, A. F., Volpi, P. S., de Souza, L. C., Comério, M., Filho, A. C., et al. (2023). Genomic-assisted breeding for climate-Smart Coffee. *Plant Genome.* 17. doi: 10.1002/tpg2.20321

Friedel, M. J., Stewart, D., Lu, X. F., Stevenson, P., Manly, H., and Dyer, T. (2023). A simple stacked ensemble machine learning model to predict naturalized catchment hydrology and allocation status. *arXiv preprint.* arXiv:2312.01649.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Ann. Stat* 19, 1–67. doi: 10.1214/aos/1176347963

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., and Suganthan, P. N. (2022). Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* 115, 105151. doi: 10.1016/j.engappai.2022.105151

Ghasemieh, A., Lloyed, A., Bahrami, P., Vajar, P., and Kashef, R. (2023). A novel machine learning model with stacking ensemble learner for predicting emergency readmission of heart-disease patients. *Decision Analytics J.* 7, 100242. doi: 10.1016/j.dajour.2023.100242

Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* 183, 347–363. doi: 10.1534/genetics.109.103952

Gokavi, N., Gangadharappa, P. M., Sathish, D., Nishani, S., Hiremath, J. S., and Koulagi, S. (2023). Genetic variability, heritability and genetic advance for quantitative traits of Arabica coffee (*Coffea Arabica* L.) genotypes. *Plant Genet. Resources: Characterization Utilization* 21, 260–268. doi: 10.1017/S1479262123000680

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction. 2ed* (New York: Springer), p. 745. doi: 10.1007/978-0-387-84858-7

Hernandez, C. O., Wyatt, L. E., and Mazourek, M. R. (2020). Genomic prediction and selection for fruit traits in winter squash. *G3 Genes|Genomes|Genetics* 10, 3601–3610. doi: 10.1534/g3.120.401215

James, G., Witten, D., Hastie, T., Tibshirani, R., and Taylor, J. (2023). *An introduction to statistical learning with applications in Python.* (Cham, Switzerland: Springer International Publishing).

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Appl. Genet.* 127, 595–607. doi: 10.1007/s00122-013-2243-1

Kalule, R., Abderrahmane, H. A., Alameri, W., and Sassi, M. (2023). Stacked Ensemble Machine Learning for porosity and absolute permeability prediction of carbonate rock plugs. *Sci. Rep.* 13. doi: 10.1038/s41598-023-36096-2

Kandel, I., Castelli, M., and Popovič, A. (2021). Comparing stacking ensemble techniques to improve musculoskeletal fracture image classification. *J. Imaging* 7, 100. doi: 10.3390/jimaging7060100

Koenker, R., and Bassett, G. (1978). Regression quantiles. *Econometrica* 46, 33. doi: 10.2307/1913643

Larkin, D. L., Lozada, D. N., and Mason, R. E. (2019). Genomic selection—considerations for successful implementation in wheat breeding programs. *Agronomy* 9, 1–18. doi: 10.3390/agronomy9090479

Lenz, P. R., Nadeau, S., Mottet, M., Perron, M., Isabel, N., Beaulieu, J., et al. (2019). Multi-trait genomic selection for Weevil Resistance, growth, and wood quality in Norway spruce. *Evolutionary Appl.* 13, 76–94. doi: 10.1111/eva.12823

Li, Z., Gao, N., Martini, J. W. R., and Simianer, H. (2019). Integrating gene expression data into genomic prediction. *Front. Genet.* 10. doi: 10.3389/fgene.2019.00126

Liang, M, Chang, T, An, B, Duan, X, Du, L, Wang, X, et al. (2021). Stacking ensemble learning framework for genomic prediction. *Front Genet.* 4, 12. doi: 10.3389/fgene.2021.600040

Liaw, A., and Wiener, M. (2002). Classification and regression by randomforest. *R J* 2, 18–22.

Long, N., Gianola, D., Rosa, G. J., and Weigel, K. A. (2011). Application of support vector regression to genome-assisted prediction of quantitative traits. *Theor. Appl. Genet.* 123, 1065–1074. doi: 10.1007/s00122-011-1648-y

McKinney, B. A., Reif, D. M., Ritchie, M. D., and Moore, J. H. (2006). Machine learning for detecting gene-gene interactions. *Appl-Bioinformatics* 5, 77–88. doi: 10.2165/00822942-200605020-00002

Meinshausen, N. (2017). *Quantregforest: Quantile regression forests - the comprehensive R.* Available online at: https://cran.r-project.org/web/packages/quantregForest/quantregForest.pdf (Accessed January 17, 2023).

Meinshausen, N. (2006). Quantile regression forests. *J. Mach. Learn. Res.* 7, 983–999.

Mendes-Moreira, J., Soares, C., Jorge, A. M., and Sousa, J. F. (2012). Ensemble approaches for regression. *ACM Computing Surveys* 45, 1–40. doi: 10.1145/2379776.2379786

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157. doi: 10.1093/genetics/157.4.1819

Mienye, I. D., and Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access* 10, 99129–99149. doi: 10.1109/ACCESS.2022.3207287

Milborrow, S. (2017).earth: multivariate adaptive regression splines. In: *R package version 4.6.3*. Available online at: https://CRAN.R-project.org/package=earth (Accessed 7 June 20123).

Mohammed, A., and Kora, R. (2023). A comprehensive review on ensemble deep learning: opportunities and challenges. *J. King Saud Univ-Comput Inf Sci.* 35, 757–774. doi: 10.1016/j.jksuci.2023.01.014

Montesinos-López, O. A., Martín-Vallejo, J., Crossa, J., Gianola, D., Hernández-Suárez, C. M., Montesinos-López, A., et al. (2019). A benchmarking between deep learning, support Vector Machine and bayesian threshold best linear unbiased prediction for predicting ordinal traits in plant breeding. *G3 Genes|Genomes|Genetics* 9, 601–618. doi: 10.1534/g3.118.200998

Montesinos López, O. A., Montesinos López, A., and Crossa, J. (2022). "Random forest for genomic prediction," in *Multivariate statistical machine learning methods for genomic prediction*. Eds. O. A. Montesinos López, A. Montesinos López and J. Crossa (Springer, Cham), pp 633–pp 681.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to linear regression analysis.* (Hoboken, New Jersey: John Wiley & Sons, Inc).

Nascimento, A. C., Nascimento, M., Azevedo, C., Silva, F., Barili, L., Vale, N., et al. (2019). Quantile regression applied to genome-enabled prediction of traits related to

flowering time in the common bean. *Agronomy* 9, 796. doi: 10.3390/agronomy9120796

Nascimento, M., Silva, E. F. F., de Resende, M. D. V., Cruz, C. D., Nascimento, A. C. C., Viana, J. M. S., et al. (2017). Regularized quantile regression applied to genome-enabled prediction of quantitative traits. *Genet. Mol. Res.* 16. doi: 10.4238/gmr16019538

Pérez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics.* 198. doi: 10.1534/genetics.114.164442

Porto, A. C. V., Freitas-Silva, O., de Souza, E. F., and Gottschalk, L. M. F. (2019). Effect of asparaginase enzyme in the reduction of asparagine in green coffee. *Beverages* 5, 32. doi: 10.3390/beverages5020032

R Core Team (2022). *R: A Language and Environment for Statistical Computing* (Vienna: R Foundation for Statistical Computing).

Seyum, E. G., Bille, N. H., Abtew, W. G., Munyengwa, N., Bell, J. M., and Cros, D. (2022). Genomic selection in tropical perennial crops and plantation trees: A review. *Mol. Breed.* 42. doi: 10.1007/s11032-022-01326-4

Sousa, T. V., Caixeta, E. T., Alkimim, E. R., Oliveira, A. C., B de Pereira, A.A., Sakiyama, N. S., et al. (2017). Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. *Tree Genet. Genomes* 13, 124. doi: 10.1007/s11295-017-1208-y

Sousa, T. V., Caixeta, E. T., Alkimim, E. R., Oliveira, A. C. B., Pereira, A. A., Sakiyama, N. S., et al. (2019). Early selection enabled by the implementation of genomic selection in coffea arabica breeding. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01934

Sousa, I. C., Nascimento, M., Silva, G. N., Nascimento, A. C. C., Cruz, C. D., Silva, F. F., et al. (2021). Genomic prediction of leaf rust resistance to Arabica coffee using machine learning algorithms. *Scientia Agricola* 78. doi: 10.1590/1678-992x-2020-0021

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91. doi: 10.3168/jds.2007-0980

Voss-Fels, K. P., Cooper, M., and Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* 132, 669–686. doi: 10.1007/s00122-018-3270-8

Weldemichael, G., Alamerew, S., and Kufa, T. (2017). Genetic variability, heritability and genetic advance for quantitative traits in coffee (Coffea arabica l.) accessions in ethiopia. *Afr. J. Agricul. Res.* 12. doi: 10.5897/AJAR2016.12059

Xu, Y., Liu, X., Fu, J., Wang, H., Wang, J., Huang, C., et al. (2019). Enhancing genetic gain through genomic selection: From livestock to plants. *Plant Commun.* 16. doi: 10.1016/j.xplc.2019.100005

Zhang, W., and Goh, A. T. C. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci. Front.* 7, 45–52. doi: 10.1016/j.gsf.2014.10.003

Zhang, J., Liu, F., Reif, J. C., and Jiang, Y. (2021). On the use of GBLUP and its extension for GWAS with additive and epistatic effects. *G3 Genes|Genomes|Genetics* 11. doi: 10.1093/g3journal/jkab122