

ARTICLE**Biometry, Modeling, and Statistics**

Corn grain yield forecasting by satellite remote sensing and machine-learning models

Antonio Alves Pinto¹  | Cristiano Zerbato¹ | Glauco de Souza Rolim¹ | Marcelo Rodrigues Barbosa Júnior¹ | Luis Fernando Vieira da Silva² | Romário Porto de Oliveira¹

¹Dep. of Rural Engineering and Exact Sciences, School of Agricultural and Veterinarian Sciences, São Paulo State Univ. (UNESP), Jaboticabal, São Paulo, Brazil

²College of Agriculture ‘Luiz de Queiroz’, Soil Science and Plant Nutrition Dep., Univ. of São Paulo, Av. Pádua Dias, 11, Piracicaba, SP 13418900, Brazil

Correspondence

Antonio Alves Pinto, Dep. of Rural Engineering and Exact Sciences, São Paulo State Univ. (UNESP), Jaboticabal, São Paulo, Brazil.

Email: antoniofca@gmail.com

Assigned to Associate Editor Nathan DeLay.

Abstract

This study aimed to evaluate the performance of six machine-learning models in forecasting corn (*Zea mays L.*) grain yield before harvest using, as input, variables in the models, some of the most-used vegetation indices (VIs) and spectral bands in the literature, as well as using data at 770 and 980 sum of degree days (SDD). The field study was carried out in a commercial area in the 2017–2018 and 2018–2019 harvests. Spectral data were obtained from Sentinel-2 satellite images and were used as input variables in the proposed models: artificial neural networks (ANN), *k*-nearest neighbors (KNN), random forest (RF), and support vector machine (SVM). The maximum R^2 and minimum values of mean absolute error (MAE) and RMSE were 0.89, 0.33, and 0.42 t ha⁻¹, respectively, for the RF algorithm using all input variables. The results obtained in the present study show that it is possible to predict corn grain yield 80 d before harvest using only VIs for the crop. Testing the various combinations of spectral bands and VIs resulted in obtaining the GREEN band and the VI global environment monitoring index (GEMI) as the best predictor variables in the present study. The use of more than one SDD did not improve the performance of the models tested. The models developed using data at 980 SDD obtained the best precision and accuracy performance both in the scenario with all model input variables and with the two best predictors. The KNN algorithm obtained the best performance in the precision and accuracy metrics for most of the scenarios studied in the present work.

Abbreviations: ANN, artificial neural network; DAS, days after sowing; GEMI, global environment monitoring index; GNDVI, green normalized difference vegetation index; KNN, *k*-nearest neighbor; MAE, mean absolute error; NDRE, normalized difference red-edge; NDVI, normalized difference vegetation index; RF, random forest; SDD, sum of degree days; SFS, ‘SequentialFeatureSelector’; SVM, support vector machine; VI, vegetation index.

© 2022 The Authors. Agronomy Journal © 2022 American Society of Agronomy.

1 | INTRODUCTION

Corn (*Zea mays* L.) is among the most cultivated cereals in the world, with more than one billion tons produced per year, most of which is produced in developing countries (in the case of Brazil), where the subsistence of the most vulnerable is dependent on the production of this crop, therefore, corn production is considered critical in supporting global food security (Leroux et al., 2019). In Brazil, corn is the second most cultivated grain, with ~29% of the area cultivated with grain in the country (19.8 million ha) in the 2020–2021 crop, producing ~93.4 million t, keeping Brazil among the largest producers and world corn exporters (Companhia Nacional de Abastecimento, 2021).

Considering the importance of corn in Brazilian agribusiness, forecasting grain yield at the field level is fundamental in decision making regarding interventions in areas with low grain yield forecasts such as fertilizer management, determination of specific management zones, harvest planning, as well as analysis of economic returns and national and global food security (Barzin et al., 2020; Venancio et al., 2019).

In recent years, with technological advances, many researchers have intensified the use of machine-learning techniques for estimating or forecasting grain yields in crops such as cotton (*Gossypium* spp.) (Haghverdi et al., 2018), wheat (*Triticum aestivum* L.) (Hunt et al., 2019; Zhou et al., 2021), sugarcane (*Saccharum officinarum* L.) (Canata et al., 2021; Maldaner et al., 2021b), and corn (Ramos et al., 2020).

Some of the most-used machine-learning models to solve complex problems such as crop grain yield forecasting are artificial neural network (ANN), *k*-nearest neighbors (KNN), random forest (RF), and support vector machines (SVM) (Ramos et al., 2020). Such models make use of an approach based on supervised learning, in which models are trained knowing the input and output variables of the models. Among the most-used input variables in crop grain yield forecasting models, climatic variables, fertility, and soil moisture stand out (Klompenburg et al., 2020). However, some studies have only evaluated the use of spectral data (Barzin et al., 2020; Canata et al., 2021; Haghverdi et al., 2018; Ramos et al., 2020), which makes it an alternative with an easy application in the field level; however, the addition of climatic variables can make the models more accurate (Tedesco, Moreira, et al., 2021).

In corn, Saranya and Nagarajan (2020) used data from the normalized difference vegetation index (NDVI) to estimate corn grain yield with the ANN and multiple linear regression models; however, studies developed by Sonobe et al. (2018), Ramos et al. (2020), and Barzin et al. (2020) reported that other vegetation indices (Vis), such as the global environment monitoring index (GEMI), green normalized difference vegetation index (GNDVI), green leaf index, normalized

Core Ideas

- A combination of different VIs (20) can improve model accuracy for corn yield forecasting.
- GREEN band and GEMI were the best predictive features in the present study.
- Models using data at 980 SDD had the best precision and accuracy performance.
- The KNN algorithm performed best in precision and accuracy metrics.

difference red-edge (NDRE), simplified canopy chlorophyll content index, and visible atmospherically resistant index, can be used in the estimation and forecasting of corn grain yield. In addition, these works estimate grain yield and not forecast. Estimation is the current quantification of a variable, while the forecast is the quantification for the future, both from historical data.

Few studies are found regarding forecast, such as the study by Kayad et al. (2019) when forecasting corn grain yield with the RF algorithm between 15 to 45 d before harvest, using GNDVI as the input variable, under controlled water conditions with center pivot irrigation in Ferrara, northern Italy. While Barzin et al. 2020, when studying some IVs in the performance of multiple regression models in the state of Mississippi, USA, were able to forecast grain yield ~75 d before harvest, using IVs simplified canopy chlorophyll content index, green leaf index, and visible atmospherically resistant index.

As stated above, some studies have already been developed seeking to estimate or predict corn grain yield; however, the forecasting of grain yield using only remote sensing data of the crop is a challenging task given that there are numerous factors that contribute to spatial variability and temporal of the crops' biophysical characteristics and their spectral data, making a model developed for a region with specific conditions not have the same performance in another region as stated by Kayad et al. (2019).

The sum of degree days (SDD) is a technique widely used in crop modeling usually to determine the duration of periods based on temperature (Aslam et al., 2017). For example, in soybean [*Glycine max* (L.) Merr.] (Akyuz et al., 2017), oil palm (*Elaeis guineensis* Jacq.) (Suresh et al., 2021), and grapevine (*Vitis vinifera* L.) (Charalampopoulos et al., 2021). Few works used the concept of degree days associated with spectral remote sensing data to model crop grain yield with machine-learning techniques (Tedesco, Moreira, et al., 2021; Tedesco, Oliveira, et al., 2021); until the moment of this study, none were found with corn cultivation.

The present study aimed to evaluate the performance of six machine-learning models in forecasting corn grain yield using as input variables of the models the combination of some of the vegetation indices and spectral bands most used in the literature obtained at 770, 980, and the combination of 770 + 980 SDD. Questions to be answered throughout our study include the following:

1. Which spectral variables should we use in forecasting grain yield, spectral bands, vegetation indices, or both?
2. Can the combination of data from more than one SSD improve the performance of the models?
3. When is the best time to collect the data?
4. Among the most-used machine-learning algorithms in problem solving, which one has the best performance?

2 | MATERIALS AND METHODS

In this section, we present information on the study areas (Section 2.1.), obtaining and processing satellite images (Section 2.2.), climatic data (Section 2.3.), obtaining grain yield data (Section 2.4.), feature selection (Section 2.5.), machine-learning algorithms (Section 2.6.), and model performance analysis (Section 2.7.).

2.1 | Study area

The study was carried out in a commercial area of 20.7 ha located in the municipality of Primavera do Oriente, Mato Grosso, at an average elevation of 636 m during the 2017–2018 and 2018–2019 harvests (Figure 1a) representing an important area corn production in Brazil (Figure 1b). The climate in the study region is classified as tropical hot and humid (Aw), which includes drought in winter (from June to September) and rains from October through April with low climate variability according to the data of precipitation and average monthly temperature from 2000 to 2021 (Figure 1c). The average annual precipitation of the historical series ranged from ~1,140 to 2,060 mm, and the average annual temperature of the historical series for the region varies between 24.1 and 25.5 °C.

The corn hybrid DKB 390 was used in both crops. The 2017–2018 crop was sown on 28 Feb 2018 and harvested with a harvester on 20 July 2018, totaling 135 d after sowing (DAS). The 2018–2019 crop was sown on 8 Feb. 2019 and harvested on 23 June 2019, totaling 142 DAS.

2.2 | Obtaining and processing satellite images

All images were obtained from the United States Geological Survey (USGS) through the Earth Explorer platform

(<https://earthexplorer.usgs.gov>). The images were corrected using the open-source Semi-Automatic Classification Plugin, v7.8.1 (Congedo, 2020) in Quantum GIS (QGIS, 2021) v3.16. This correction consists of converting the digital numbers into reflectance data based on the Dark Object Subtraction atmospheric correction methodology (Chavez, 1996).

After correction, the images were cropped to extract the area of interest. In obtaining the images, we identified that the images corresponding to the beginning of the crop development for both seasons had a problem with cloud cover (Wang et al., 2019), thus, four Sentinel-2 images with low cloud cover (<1%) were used, two in the first and two in the second crop at 55 and 70 DAS, which corresponds to the reproductive phases V10 and R1, respectively; these phenological phases correspond to 770 and 980 SDD. These phenological phases have already been used as a reference to obtain spectral data from corn that were later used in grain yield estimation models (Barzin et al., 2020; Kayad et al., 2019; Ramos et al., 2020).

In the study, four Sentinel-2 spectral bands were considered to compose part of the model's input variables (Table 1).

In addition to the spectral bands, 16 VIs were obtained from the images (Table 2), these indices are cited in the literature for monitoring grain yields of different crops (Barzin et al., 2020; Canata et al., 2021; Haghverdi et al., 2018; Ramos et al., 2020; Zhou et al., 2021).

2.3 | Climate data

To determine the SDD in the two corn crops, we used data from a meteorological station located ~15 km from the production fields. For each crop, we collected the daily maximum and minimum temperature data from the sowing date to the harvest. The calculation was performed according to the acquisition date of each satellite image following the Arnold method (1959):

$$\text{SSD} = \sum \left(\frac{T_{\max} + T_{\min}}{2} \right) - T_{\text{base}} \quad (1)$$

where SDD is the sum of degree-days, T_{\max} is the maximum air temperature (°C), T_{\min} is the minimum air temperature (°C), and T_{base} is the lower baseline temperature of the crop (10 °C). For images at 55 DAS, the SDD was 770 ± 3 (1,030 ± 5 SDD before harvest), and for 70 DAS, it was 980 ± 3 SDD (820 ± 5 SDD before harvest).

After obtaining the SDD of each image, we organized the data (spectral bands and VIs), to develop models with data at 770 SDD, at 980 SDD, and the merging of data, 770 + 980 SDD.

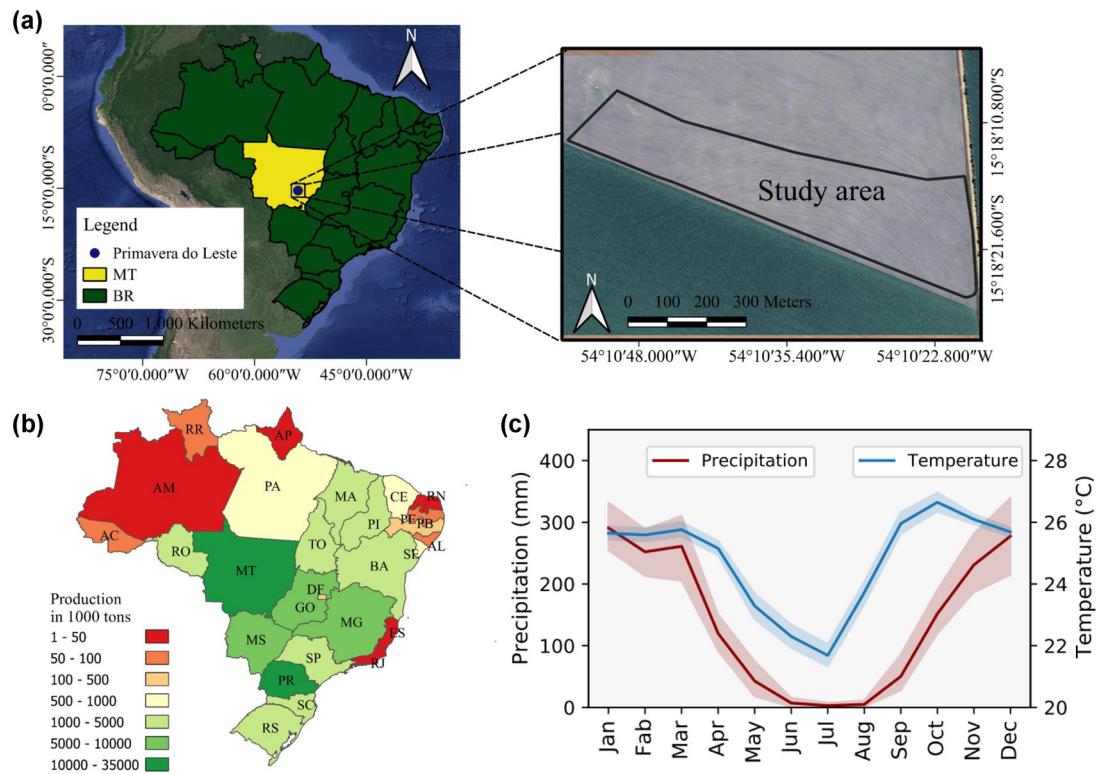


FIGURE 1 (a) Location of the study area in Brazil. (b) Estimated corn production in Brazil in the 2020–2021 crop. Adapted from Companhia Nacional de Abastecimento, 2021. (c) Monthly precipitation and temperature from 2000 to 2021. ANN, artificial neural networks; KNN, *k*-nearest neighbors; RF, random forest; SVM, support vector machines; SDD, sum of degree days; NDVI, normalized difference vegetation index

TABLE 1 Specifications of the spectral bands from Sentinel-2

Spectral bands	Central wavelength	Bandwidth	Resolution		
			Spatial	Temporal	Radiometric
Blue	492.4	66	m	d	bits
Green	559.8	36			
Red	664.6	31			
Near-infrared	832.8	106			

2.4 | Obtaining grain yield data

The grain yield data observed for corn in both seasons were obtained by sensors embedded in the harvester that collected ~600 georeferenced grain yield points per hectare. The sensors were calibrated before the harvest process to minimize measurement errors in the field. After collection, the data were processed by removing the outliers identified in the processing according to the methodology described by Maldane et al. (2021a), which takes into account the local and global variability in data processing. After processing, grain yield maps of the two crops were generated, allowing such data to be used for training and testing the models. For each crop year,

Pearson correlations between grain yield and features were estimated.

2.5 | Feature selection

In this research, we used the ‘SequentialFeatureSelector’ (SFS) algorithm from the Python 3.6 scikit-learn library to select the most important variables among all spectral bands (Table 1) and IVs (Table 2). The algorithm works in stages, iteratively adding features forming a subset of selected features, that is, at each stage, this algorithm chooses the best feature to add based on the cross-validation score of an

TABLE 2 Vegetation indices (Vis) used in this study

VI	Equation	Reference
ARVI	$(\text{NIR} - [\text{Red} - 1(\text{Red} - \text{Blue})]) / (\text{NIR} + [\text{Red} - 1(\text{Red} - \text{Blue})])$	Kaufman and Tanré (1992)
EVI	$(2.5)(\text{NIR} - \text{Red}) / (\text{NIR} + 6[\text{Red} - 7.5(\text{Blue} + 1)])$	Huete et al. (1997)
EVI2	$2.5[(\text{NIR} - \text{Red}) / (\text{NIR} + 2.4\text{Red} + 1)]$	Zhangyan et al. (2008)
GARI	$\text{NIR} - [\text{Green} - (\text{Blue} - \text{Red})] / (\text{NIR} + [\text{Green} - (\text{Blue} - \text{Red})])$	Gitelson et al. (1996)
GDVI	$\text{NIR} - \text{Green}$	Wu (2014)
GEMI	$\eta(1 - 0.25\eta) - [(\text{Red} - 0.125) / (1 - \text{Red})]^a$	Pinty and Verstraete (1991)
GLI	$(\text{Green} - \text{Red} - \text{Blue}) / (2\text{Green} + \text{Red} + \text{Blue})$	Louhaichi et al. (2001)
GNDVI	$(\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green})$	Gitelson et al. (1996)
GSAVI	$1.5(\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green} + 0.5)$	Sripada et al. (2008)
MNLI	$(\text{NIR}^2 - \text{Red})(1.5) / (\text{NIR}^2 + \text{Red} + 0.5)$	Gong et al. (2003)
NLI	$(\text{NIR}^2 - \text{Red}) / (\text{NIR}^2 + \text{Red})$	Goel and Qin. (2009)
NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	Rouse et al. (1974)
OSAVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 0.16)$	Rondeaux et al. (1996)
SAVI	$[(\text{NIR} - \text{Red}) (1 + 0.5)] / [(\text{NIR} + \text{Red}) + 0.5]$	Huete (1988)
VARI	$(\text{Green} - \text{Red}) / (\text{Green} + \text{Red} - \text{Blue})$	Gitelson et al. (2002)
WDRVI	$(0.1\text{NIR} - \text{Red}) / (0.1\text{NIR} + \text{Red})$	Gitelson (2004)

Note. NIR, near infrared; ARVI, atmospherically resistant vegetation index; EVI, enhanced vegetation index; EVI2, enhanced vegetation index 2; GARI, green atmospherically resistant vegetation index; GDVI, green difference vegetation index; GEMI, global environment monitoring index; GLI, green leaf index; GNDVI, green normalized difference vegetation index; GSAVI, green soil-adjusted vegetation index; MNLI, modified non-linear index; NLI, nonlinear index; NDVI, normalized difference vegetation index; OSAVI, optimized soil-adjusted vegetation index; SAVI, soil-adjusted vegetation index; VARI, visible atmospherically resistant index; WDRVI, wide dynamic range vegetation index.

^a $\eta = [2(\text{NIR}^2 - \text{RED}^2) + 1.5\text{NIR} + 0.5\text{RED}] / (\text{NIR} + \text{RED} + 0.5)$.

estimator. For this study, we used the RF algorithm as the SFS estimator. Random forest is one of the most-used machine-learning methods in data science works, achieving good performance in the selection of variables (Barzin et al., 2020; Ramos et al., 2020). In the selection of features, we used data from the two SDDs (770 and 980) from the two seasons. Before selecting features and training the models, all predictor variables were standardized.

2.6 | Machine-learning algorithms

In the present study, the machine-learning models used were ANN, KNN, RF, and SVM with their respective algorithms, MLPRegressor, KNeighborsRegressor, RandomForestRegressor, and SVR from the scikit-learn library of Python language.

The ANN is a computational structure capable of acquiring and storing knowledge through learning mechanisms, allowing the resolution of complex problems (Cherukuri et al., 2019). Neural networks consist of an input layer, one or more hidden layers, and an output layer; however, the number of hidden layers can influence the time and performance of the model (Zhou et al., 2021). Thus, in the study, we used multilayer perceptron-type ANN with one, two, and three hidden layers corresponding to ANN-1,

ANN-2, and ANN-3, respectively, rectified linear activation function, maximum number of iterations of 5,000, learning rate of 0.0001, initial learning rate of type ‘adaptive’, and the weight optimization solver was the Broyden–Fletcher–Goldfarb–Shanno limited-memory optimization algorithm, which is an optimizer of the family of methods quasi-Newton, the number of neurons tested per layer ranged from two to 20, with eight being selected for ANN-1, four and five neuron for ANN-2, and five, five, and three for ANN-3.

The KNN algorithm is used in classification and regression problems. The algorithm takes into account the nearest neighbors to define the weights so that the closest points contribute more to the average of the predicted values (Abbas et al., 2020). In KNN, k is an adjusted parameter with an important role in obtaining an accurate forecast; in the study, k was used equal to 16, and distance from the neighbor was calculated using the Minkowski distance.

The RF algorithm is a learning technique that makes use of a set of decision trees, with each tree being dependent on the values of a random vector (Maldaner et al., 2021b). In RF, the data are recursively divided into more homogeneous units in each tree (nodes), thus, the predicted corn grain yield values are the average of the adjusted response of all the individual trees originating from each initialized sample. To obtain good results in the RF, the number of trees (ntree) was adjusted to

300 trees and the number of variables available for division in each node of the tree (*mtry*) to five.

Support vector machine is one of the most popular machine-learning techniques for classification and regression (Zhou et al., 2021). The algorithm seeks to minimize the error by adding to the hyperplane and maximize the margin between predicted and observed values. In the SVM used in the study, it was necessary to define three parameters: the kernel function using the radial basis function, the Kernel coefficient for the radial basis function, which was one divided by the number of independent variables and the equal loss function to 0.1.

The selection of the best fit of the parameters for each model was through the exhaustive search algorithm GridSearchCV from the scikit-learn library. This algorithm tests all combinations of the parameters passed in an orderly way and seeking the general minimum of the sum of squares function between the data observed and predicted by the models. The GridSearchCV allows performing cross-validation and selecting one or more evaluation metrics in the selection of the best parameters.

2.7 | Model performance analysis

In the research, cross-validation was used to evaluate the performance of the models. We use the KFold package from the Scikit-learn library, where k represents the number of divisions of the total dataset into smaller subsets. For this research, we used the value of $k = 10$, dividing the data into 10 smaller subsets at random so that for each execution, a subset was used to evaluate the estimated results from the other nine, this process being repeated 10 times until all 10 subsets were used once in validation (Xu et al., 2019). After the tests, the mean values of the cross-validation were obtained for the coefficient of determination (R^2), the mean absolute error (MAE) and the root mean error square (RMSE). The synthesis of the methodology is presented in Figure 2. With the results of the metrics for each model, the Scott-Knott means test was performed for comparing the performance of the tested multilayer models.

3 | RESULTS

For each crop year, the Pearson correlation of grain yield with the 16 VIs and four spectral bands at 770 and 980 SDD is presented (Figure 3a,b). In this sense, at 770 SDD, the highest positive correlation with grain yield in both crops was for GNDVI with 0.51 in the 2017–2018 crop, while in the 2018–2019 crop it was 0.49. In the data at 980 SDD, the highest positive correlation was GARI (0.74) in the 2017–2018 harvest and GNDVI (0.68) in the 2018–2019 harvest. For the

variables negatively correlated with grain yield, the GREEN band was superior to the others with -0.49 and -0.55 at 770 SDD in the 2017–2018 and 2018–2019 harvests, respectively.

We also emphasize that in both seasons the correlation of grain yield with the input variables was higher for the data at 980 SDD (Figure 3b), in which, with the exception of VARI in the 2018–2019 season, all variables had a positive correlation or negative >0.4 .

The increase in R^2 (Figure 4) and decrease in RMSE stabilized with three variables (GREEN, GEMI, and RED) with precision (R^2) equal to .83 and model accuracy (RMSE) equal to 0.52 t ha^{-1} , with an increase of 48 and 40%, respectively, in relation to the model with one input variable. Although, with three variables, the model has better precision and accuracy; this improvement is close to the values obtained when the model was developed with the two best variables, increasing the precision and accuracy by 46 and 37%, respectively (R^2 of .82 and RMSE of 0.54 t ha^{-1}) when compared with the model with one input variable.

Considering that the increase in the number of input variables results in greater difficulty in using the models (Figure 4), we used two variables selected by the SFS algorithm for the forecasting of grain yield, these being the GREEN spectral band and the GEMI VI. In Supplemental Table S1, it is possible to check all the features selected for the number of variables from one to 20.

In terms of precision, all models developed with the spectral data at 980 SDD were superior to the other data for the models with all variables and with the two best ones (Figure 5). While in the data at 770 and 770 + 980 SDD, the averages of R^2 obtained similar values in all models evaluated with all variables and with the two best ones; however, for the data at 770 SDD, greater variability of precision was observed in relation to the 770 + 980 SDD dataset.

Regarding accuracy metrics, all algorithms evaluated with data at 980 SDD obtained lower MAE and RMSE when compared with data at 770 and 770 + 980 SDD in the scenario with all variables and with only the two best variables. As with the precision metric, most models with data of 770 + 980 SDD have less accuracy variability with values closer to the mean in each model.

It is important to mention that with the use of the two best variables, the average values of precision and accuracy of the models were lower than the models with all variables, with R^2 averages ranging between .81 and .87, MAE between 0.35 and 0.44 t ha^{-1} , and RMSE from 0.47 to 0.56 t ha^{-1} .

For the models developed with all variables, KNN and RF did not differ statistically (Table 3), obtaining the highest precision values for R^2 of .87 and .89 for the scenario with data of 770 and 980 SDD, respectively. While, for the accuracy metrics, no statistical difference was observed.

When analyzing the results of the models developed in the scenario with data of 770 and 980 SDD with all variables, we

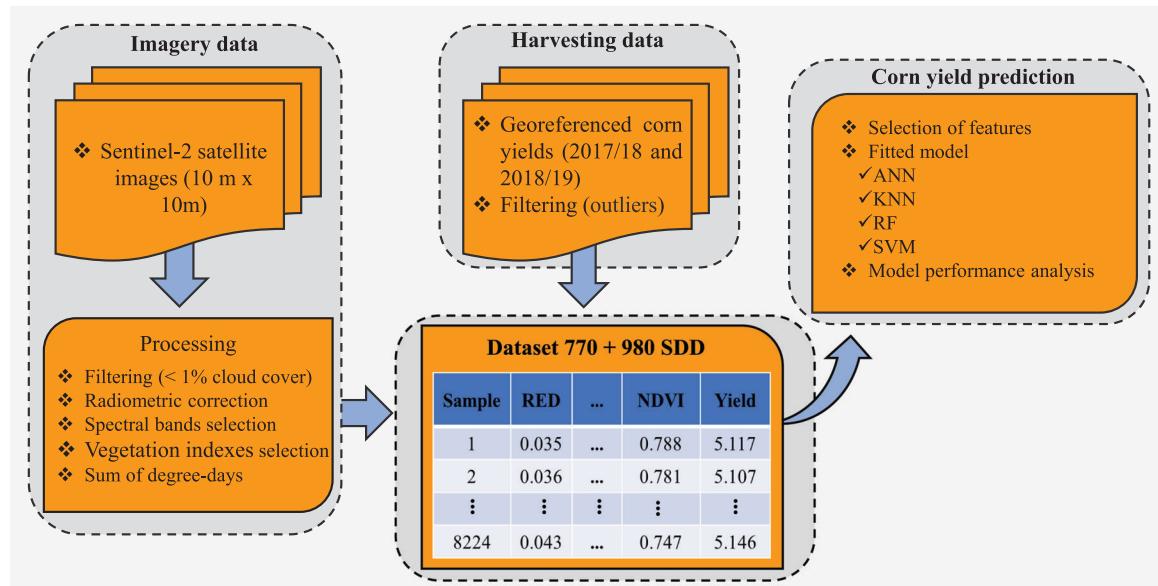


FIGURE 2 Flow chart of the methodology for data acquisition, processing, and corn grain yield forecasting

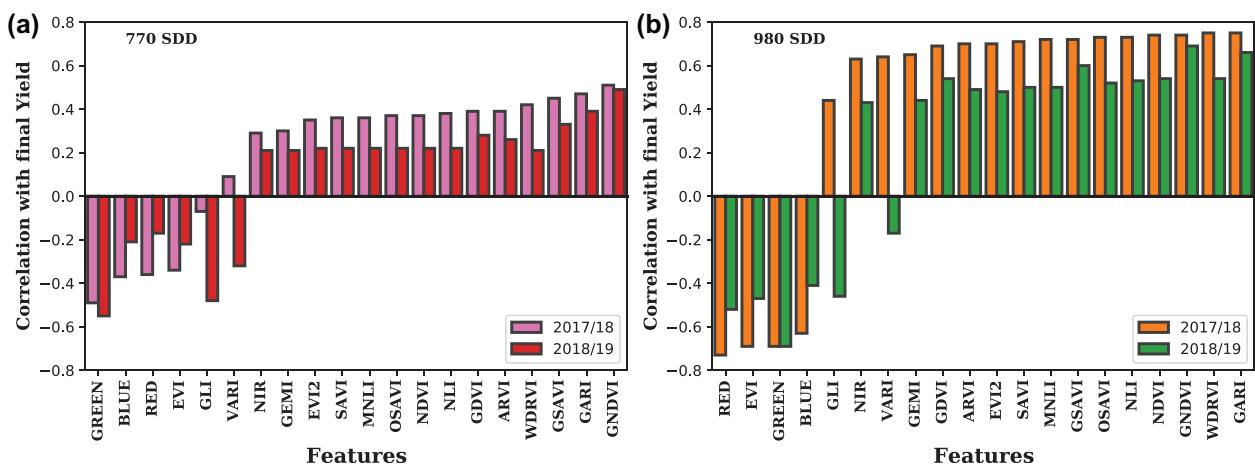


FIGURE 3 Correlation between predictor variables and grain yield. ARVI; atmospherically resistant vegetation index; EVI, enhanced vegetation index; EVI2, enhanced vegetation index 2; GARI, green atmospherically resistant vegetation index; GDVI, green difference vegetation index; GEMI, global environment monitoring index; GLI, green leaf index; GNDVI, green normalized difference vegetation index; GSAVI, green soil-adjusted vegetation index; MNLI; modified non-linear index; NLI, nonlinear index; NDVI, normalized difference vegetation index; OSAVI, optimized soil-adjusted vegetation index; SAVI, soil-adjusted vegetation index; VARI, visible atmospherically resistant index; WDRVI, wide dynamic range vegetation index

observed that the KNN algorithm obtained the best precision and accuracy, statistically differing from the other evaluated models, with averages of R^2 of .87, MAE of 0.35 t ha^{-1} , and RMSE of 0.47 t ha^{-1} .

For the models developed with the two best variables, the SVM statistically differed from the other models, obtaining the lowest precision value with R^2 of .81 in the 770 SDD data. In the 980 SDD data, the KNN and RF models were superior to the other models with $R^2 = .87$ and $\text{RMSE} = 0.47 \text{ t ha}^{-1}$. In

the scenario with data of 770 + 980 SDD, the KNN model differed from the other models with better averages of R^2 , MAE, and RMSE (Table 3).

In the data dispersion of 770 SDD with all variables (Figure 6a), the model with the best precision and accuracy was the KNN with R^2 of .87, MAE of 0.36 t ha^{-1} , and RMSE of 0.46 t ha^{-1} , with the highest forecast error frequency for errors $<0.5 \text{ t ha}^{-1}$, as well as for data from 770 SDD with the two best variables (Figure 6b), in which KNN obtained

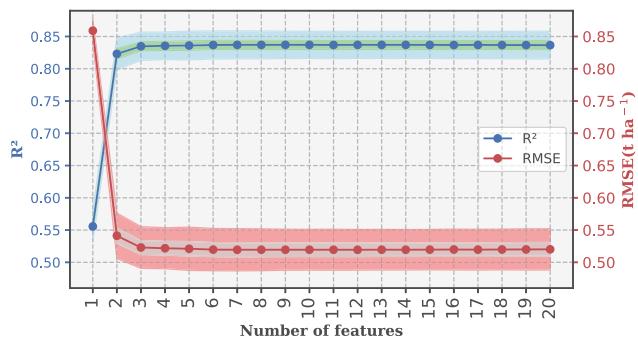


FIGURE 4 Feature importance of the model's performance

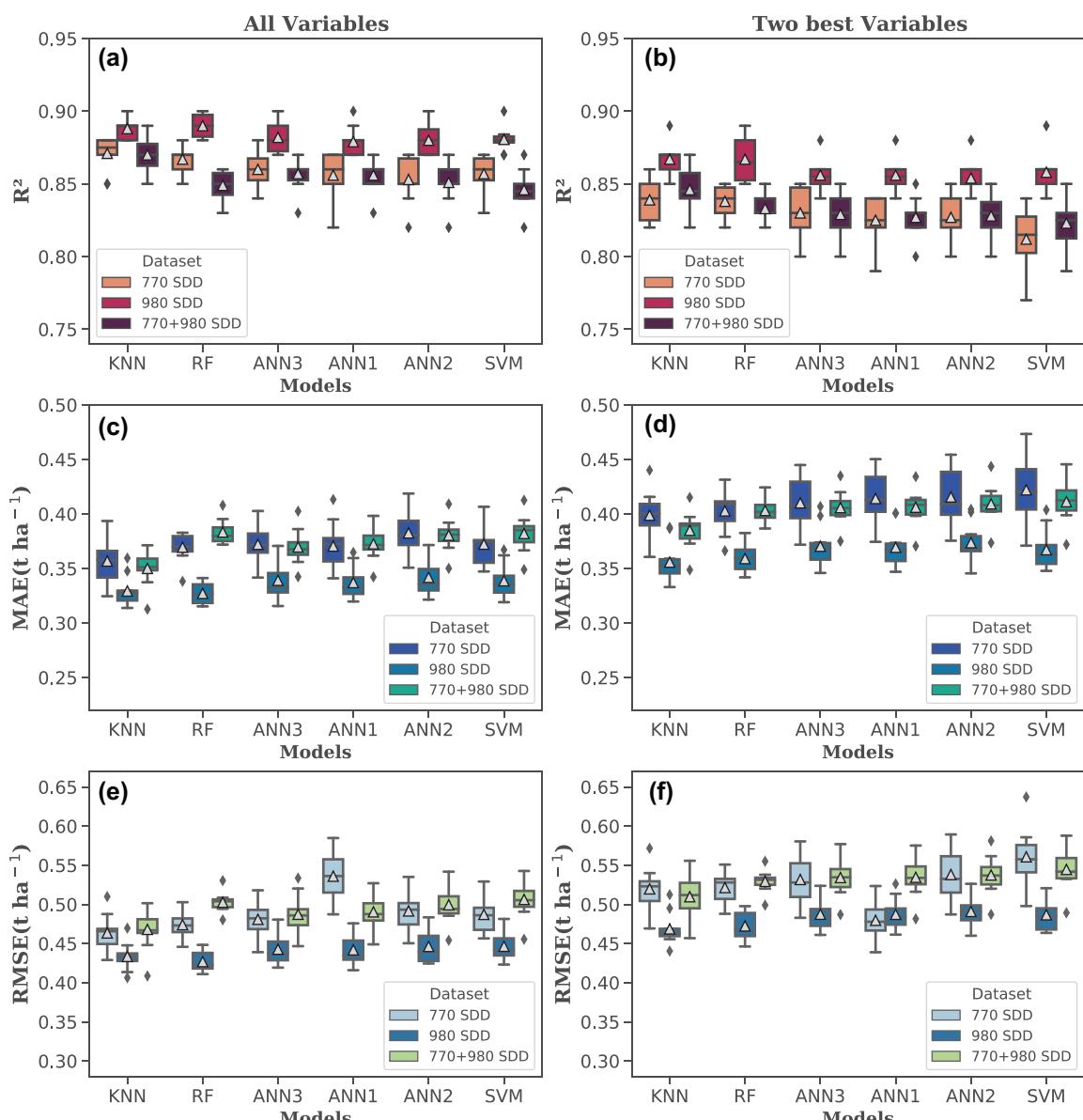


FIGURE 5 Performance metrics of each model with all the variables and the two best variables in the data with 770, 980, and 770 + 980 days after sowing. (a) R^2 all variables; (b) R^2 two best variables; (c) mean absolute error (MAE) all variables; (d) MAE two best variables; (e) RMSE all variables; (f) RMSE two best variables. ANN, artificial neural networks; KNN, *k*-nearest neighbors; RF, random forest; SVM, support vector machines; SDD, sum of degree days

better precision and accuracy with R^2 of .83, MAE of 0.40 t ha⁻¹, and RMSE of 0.51 t ha⁻¹, with most forecast errors <0.55 t ha⁻¹.

For data dispersion of 980 SDD with all variables (Figure 6c), the RF model was superior to the others with better precision (R^2 of .89) and accuracy (MAE of 0.33 t ha⁻¹ and RMSE of 0.42 t ha⁻¹), with a higher frequency of forecast error for errors <0.4 t ha⁻¹, while for the data from 70 DAS with the two best variables (Figure 6d), the KNN obtained better precision and accuracy among the models evaluated with R^2 of .87, MAE of 0.35 t ha⁻¹, and RMSE of 0.46 t ha⁻¹.

TABLE 3 Grouping of means by the Scott–Knott test for coefficients of determination (R^2), mean absolute error (MAE), and RMSE obtained with machine-learning models using all variables and the two best variables

Models	All Variables			770 SDD			980 SDD			770 + 980 SDD		
	R^2	MAE	RMSE	770 SDD			R^2	MAE	RMSE	R^2	MAE	RMSE
				770 SDD	980 SDD	770 + 980 SDD	R^2	MAE	RMSE	R^2	MAE	RMSE
All variables												
ANN-1	.85 b	0.37 a	0.49 a	.88 b	0.34 a	0.45 a	.85 b	0.37 a	0.49 a	.85 b	0.38 a	0.49 a
ANN-2	.85 b	0.38 a	0.49 a	.88 b	0.34 a	0.45 a	.85 b	0.38 a	0.49 a	.85 b	0.37 a	0.49 a
ANN-3	.85 b	0.37 a	0.48 a	.88 b	0.34 a	0.44 a	.85 b	0.37 a	0.49 a	.85 b	0.35 b	0.47 b
KNN	.87 a	0.36 a	0.46 a	.89 a	0.34 a	0.43 a	.87 a	0.35 a	0.47 b	.84 b	0.39 a	0.50 a
RF	.87 a	0.37 a	0.47 a	.89 a	0.33 a	0.43 a	.84 b	0.38 a	0.50 a	.84 b	0.38 a	0.50 a
SVM	.85 b	0.37 a	0.49 a	.87 b	0.34 a	0.45 a	.84 b	0.38 a	0.50 a	.84 b	0.38 a	0.50 a
Two best variables												
ANN-1	.82 a	0.41 a	0.53 a	.85 b	0.37 a	0.49 a	.83 b	0.41 a	0.54 a	.83 b	0.41 a	0.54 a
ANN-2	.82 a	0.42 a	0.54 a	.85 b	0.37 a	0.49 a	.83 b	0.41 a	0.54 a	.83 b	0.40 a	0.53 a
ANN-3	.83 a	0.41 a	0.53 a	.85 b	0.37 a	0.49 a	.83 b	0.40 a	0.53 a	.83 b	0.38 b	0.50 b
KNN	.83 a	0.39 a	0.52 a	.87 a	0.35 a	0.47 b	.85 a	0.38 a	0.52 a	.83 b	0.40 a	0.52 a
RF	.83 a	0.40 a	0.52 a	.87 a	0.36 a	0.47 b	.83 b	0.40 a	0.52 a	.83 b	0.44 a	0.54 a
SVM	.81 b	0.42 a	0.56 a	.85 b	0.36 a	0.48 a	.82 b	0.44 a	0.54 a	.82 b	0.44 a	0.54 a

Note. ANN, artificial neural networks; KNN, k-nearest neighbors; RF, random forest; SVM, support vector machines; SDD, sum of degree days. Means followed by different letters in the same column differ by the Scott–Knott test at 5% probability.

In the data dispersion of 770 + 980 SDD with all variables and the two best input variables of the model (Figure 6e,f), the KNN model performed better with R^2 of .87 and .85, MAE of 0.35 and 0.38 t ha⁻¹, and RMSE of 0.47 and 0.51 t ha⁻¹ using all input variables and the two best ones, respectively.

4 | DISCUSSION

The variables most positively or negatively correlated with grain yield present the spectral band of green as a common characteristic between them (Figure 3a,b). In Sentinel-2, the green band corresponds to B3 with central wavelength of 559.8 nm (Table 1). This band is characterized by high reflectance of electromagnetic radiation when compared with other spectral bands in the visible region. Accordingly, Gitelson et al. (2002) the reflectance between 520 and 630 nm is the most sensitive for chlorophyll, thus the IVs GNDVI, GARI, and the GREEN band are more sensitive to the concentration of chlorophyll in a wide range of chlorophyll variations, thus contributing, for these variables, to better performance than others, such as the NDVI, corroborating what was observed by Schwalbert et al. (2018) and Kayad et al. (2019) when they reported that some indices constituted by the green band performed well in models of corn grain yield estimation.

The highest correlations of grain yield with the input variables for the data at 980 SDD (Figure 3b) corroborate what was observed by Barzin et al. (2020) when reporting similar behavior. This behavior coincides with the change of phase

in the corn crop in each of the SDD because at 770 SDD (55 DAS) the plants were in the V10 vegetative stage, which is characterized by starting a rapid and continuous growth with an accumulation of nutrients and dry weight; moreover, at this stage only the number of rows of grains is defined while at 980 SDD (70 DAS) the plants have already stopped accumulating nutrients and dry weight, the number of ovules (potential grains) and size of the ears have already been defined, and the plants begin the fertilization of the ovules and later the filling of the grains, thus changing the drain source from the leaves to the ears (Silva et al. 2021).

Considering the variability of the correlation between the spectral data and grain yield over the two crops and the data collection period (Figure 3), we emphasize that the use of more than one VI or spectral band in the forecasting of grain yield of the corn allows the evaluation of different data configurations in the grain yield forecasting, thus increasing the possibility of obtaining better results for the evaluated models as also indicated by Ramos et al. (2020).

The selection of the two best features resulted in the selection of the GREEN spectral band and the GEMI vegetation index (Figure 4). The GEMI is an atmospherically corrected index that seeks to minimize atmospheric effects (Pinty & Verstraete, 1991). In addition, it is less sensitive to ground noise and is not affected by atmospheric disturbances. According to Madasa et al. (2021), the GEMI tends to have similar results to the NDVI because both use the same bands (red and near-infrared); however, the GEMI is not sensitive to atmospheric disturbances unlike the NDVI. This fact is related

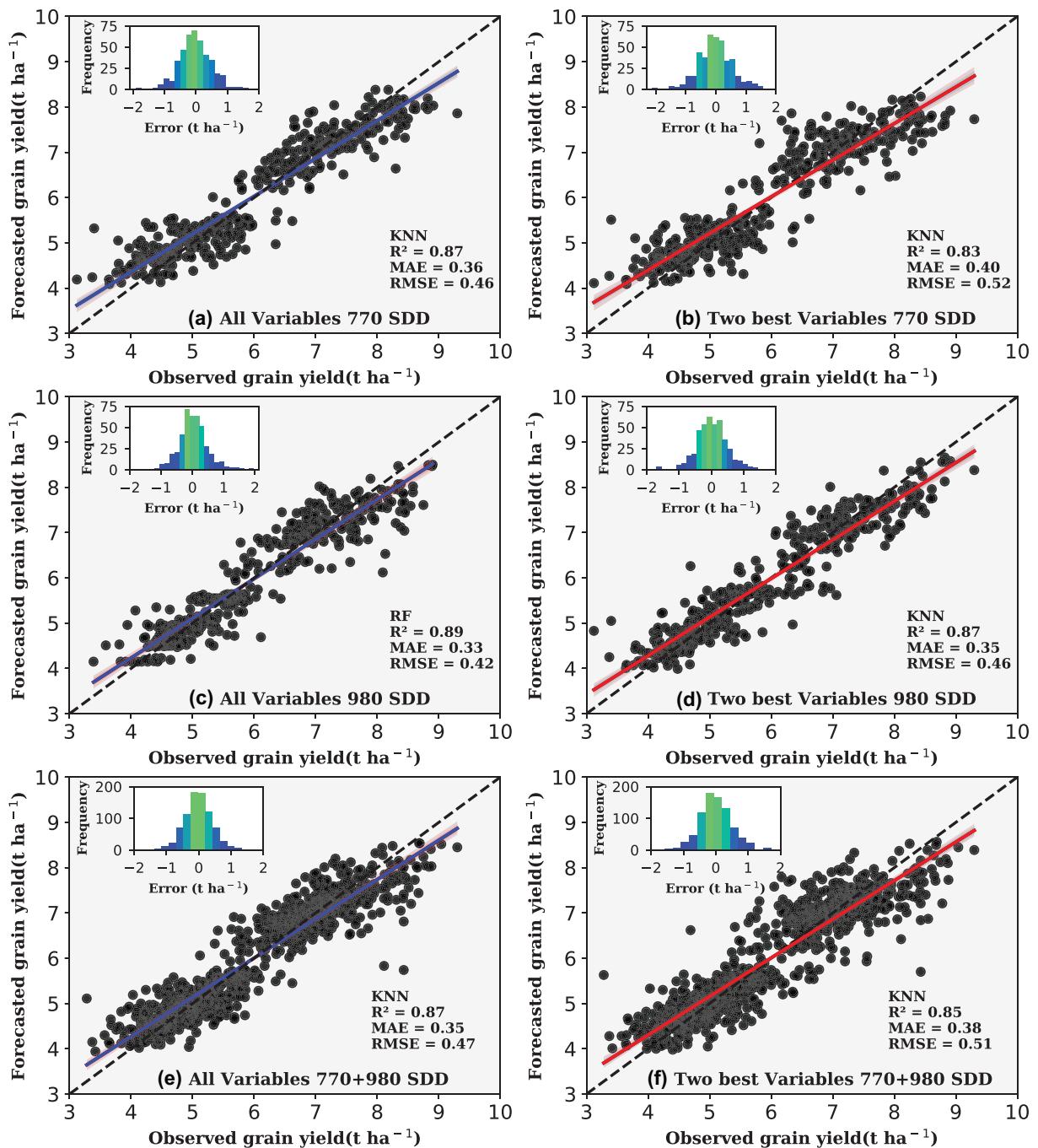


FIGURE 6 Relationships between observed and forecast grain yield obtained with machine-learning models using all variables and the two best variables. KNN, *k*-nearest neighbors; MAE, mean absolute error; SDD, sum of degree days

to the selection of the index by the algorithm, corroborating what was observed by Sonobe et al. (2018) when analyzing the performance of three algorithms (RF, SVM, and superlearner) in the classification of six crops—common bean (*Phaseolus vulgaris* L.), beetroot (*Beta vulgaris* L. subsp. *vulgaris*), grass (syn. *Urochloa*), corn, potato (*Solanum tuberosum* L.), and wheat—reported that the GEMI obtained >10% of importance in the performance of the RF for beetroot and was five times more important in SVM performance for common bean.

When studying the performance of machine-learning algorithms in forecasting corn grain yields, Khanal et al. (2018) observed that on a scale from 0 to 100 of the importance of variables for the artificial neural network model, the near-infrared, GREEN, BLUE, and RED bands obtained importance of 100, 70, 60, and 58, respectively, contributing to improving model performance.

The lower values of the precision and accuracy metrics of the models with the two best variables in relation to the

models with all the variables (Figure 5b,d,f; Table 3) corroborate what was observed by Ramos et al. (2020) when reporting lower metric values for the model with the three best features selected when compared with the model with all features for the SVM and linear regression algorithms. In addition, the precision and accuracy values of the models with the two best features of our work are superior to the results obtained by the authors when evaluating seven machine-learning models in the forecasting of corn grain yield, who obtained better results for the combined RF algorithm with regression with R^2 of .78 and MAE of 0.85 t ha⁻¹ using as input variables the three best predictors selected in the study (NDVI, NDRE, and GNDV).

The results obtained in the present study show that it is possible to forecast corn grain yield 80 d before harvest using only spectral data of the crop. These results enable improvements in corn cultivation through better management of the crop. In this sense, farmers, researchers, and government agencies can use high-resolution satellite images to monitor corn crop grain yield. This strategy can support decision making in harvesting operations such as delimiting management zones for harvesting based on grain yield forecasts, allowing harvesting operations to start in areas that have higher grain yield forecasts (smart harvesting), as well as after harvest in aiding in the logistics of transport, storage, and commercialization of grains.

Our study obtained promising results; however, we highlight the need for further studies using data from other locations with different climatic conditions, soil types, and cultivars, thus seeking to develop more robust models that can achieve good performance in forecasting crop grain yield in the most diverse conditions possible in order to assist appropriate crop management, decision making in the field of precision agriculture, as well as in controlling the stability of national and global food security.

5 | CONCLUSIONS

Testing the various combinations of spectral bands and vegetation indices resulted in obtaining the GREEN band and the GEMI band as the best predictor variables in the present study. The use of more than one SDD did not improve the performance of the models tested in our study. The models developed using data at 980 SDD obtained the best precision and accuracy performance both in the scenario with all model input variables and with the two best predictors. The KNN algorithm obtained the best performance in the precision and accuracy metrics for most of the scenarios studied in the present work.

ACKNOWLEDGMENTS

The authors are grateful to the São Paulo State University (UNESP), Jaboticabal Campus, for the academic support and

availability of laboratories, the company VSL consultoria for providing the grain yield data and the Coordination for the Improvement of Higher Education Personnel (CAPES) for the scholarship granted to the first author (Financing Code 001).

AUTHOR CONTRIBUTIONS

Antonio Alves Pinto: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Writing – original draft; Writing – review & editing. Cristiano Zerbato: Data curation; Formal analysis; Investigation; Methodology; Supervision; Writing – original draft. Glauco de Souza Rolim: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Supervision; Validation; Visualization. Marcelo Rodrigues Barbosa Júnior: Formal analysis; Supervision; Validation; Visualization; Writing – review & editing. Luis Fernando Vieira da Silva: Conceptualization; Methodology; Supervision; Writing – review & editing. Romário Porto de Oliveira: Conceptualization; Investigation; Supervision.

CONFLICT OF INTEREST

The authors declare no conflicts of interest

ORCID

Antonio Alves Pinto  <https://orcid.org/0000-0001-8615-2387>

REFERENCES

- Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, 10, 1046. <https://doi.org/10.3390/agronomy10071046>
- Akyuz, F. A., Kandel, H., & Morlocke, D. (2017). Developing a growing degree day model for North Dakota and Northern Minnesota soybean. *Agricultural and Forest Meteorology*, 239, 134–140. <https://doi.org/10.1016/j.agrformet.2017.02.027>
- Aslam, M. A., Ahmed, M., Stöckle, C. O., Higgins, S. S., Hassan, F. U., & Hayat, R. (2017). Can growing degree days and photoperiod predict spring wheat phenology? *Frontiers in Environmental Science*, 10, 57. <https://doi.org/10.3389/fenvs.2017.00057>
- Arnold, C. Y. (1959). The determination and significance of the base temperature in a linear heat unit system. *Proceedings of the American Society for Horticultural Science*, 74, 430–450.
- Barzin, R., Pathak, R., Lotfi, H., Varco, J., & Bora, G. C. (2020). Use of UAS multispectral imagery at different physiological stages for yield prediction and input resource optimization in corn. *Remote Sensing*, 12, 2392. <https://doi.org/10.3390/rs12152392>
- Canata, T. F., Wei, M. C. F., Maldaner, L. F., & Molin, J. P. (2021). Sugarcane yield mapping using high-resolution imagery data and machine learning technique. *Remote Sensing*, 13, 232. <https://doi.org/10.3390/rs13020232>
- Charalampopoulos, I., Polychroni, I., Psomiadis, E., & Nastos, P. (2021). Spatiotemporal estimation of the olive and vine cultivations' growing degree days in the Balkans region. *Atmosphere*, 12, 148. <https://doi.org/10.3390/atmos12020148>

- Chavez, P. Jr. (1996). Image-based atmospheric corrections revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62, 1025–1036.
- Cherukuri, H., Perez-Bernabeu, E., Selles, M., & Schmitz, T. (2019). Machining chatter prediction using a data learning model. *Journal of Manufacturing and Materials Processing*, 3, 45. <https://doi.org/10.3390/jmmp3020045>
- Companhia Nacional de Abastecimento. (2021). *Acompanhamento da safra brasileira grãos safra 2020/21. 7,10*. <https://www.conab.gov.br/info-agro/safras>
- Congedo, L. (2020). *Semi-automatic classification plugin documentation* (Release 6.0.1.1). <https://doi.org/10.13140/RG.2.2.29474.02242/1>
- Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology*, 161, 165–173. <https://doi.org/10.1078/0176-1617-01176>
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58, 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)
- Gitelson, A. A., Stark, R., Grits, U., Rundquist, D., Kaufman, Y., & Derry, D. (2002). Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction. *International Journal of Remote Sensing*, 23, 2537–2562. <https://doi.org/10.1080/01431160110107806>
- Goel, N. S., & Qin, W. (2009). Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. *Remote Sensing Review*, 10, 309–347. <https://doi.org/10.1080/02757259409532252>
- Gong, P., Pu, R., Biging, G. S., & Larrieu, M. R. (2003). Estimation of forest leaf area index using vegetation indices derived from hyperion hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 1355–1362. <https://doi.org/10.1109/TGRS.2003.812910>
- Haghverdi, A., Allen, R. A. W., & Leib, B. G. (2018). Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. *Computers and Electronics in Agriculture*, 152, 186–197. <https://doi.org/10.1016/j.compag.2018.07.021>
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Huete, A. R., Liu, H. Q., Batchily, K., & Leeuwen, W. V. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing Environment*, 59, 440–451. [https://doi.org/10.1016/S0034-4257\(96\)00112-5](https://doi.org/10.1016/S0034-4257(96)00112-5)
- Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, 233, 111410. <https://doi.org/10.1016/j.rse.2019.111410>
- Leroux, L., Castets, M., Baron, C., Escorihuela, M. J., Bégué, A., & Seen, D. L. (2019). Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. *European Journal of Agronomy*, 108, 11–26. <https://doi.org/10.1016/j.eja.2019.04.007>
- Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16, 65–70. <https://doi.org/10.1080/10106040108542184>
- Kayad, A., Sozzi, M., Gatto, S., Marinello, F., & Pirotti, F. (2019). Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. *Remote Sensing*, 11, 2873. <https://doi.org/10.3390/rs11232873>
- Kaufman, Y. J., & Tanré, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 261–270. <https://doi.org/10.1109/36.134076>
- Klompenburg, T. V., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. <https://doi.org/10.1016/j.compag.2020.105709>
- Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., & Shearer, S. (2018). Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture*, 153, 213–225. <https://doi.org/10.1016/j.compag.2018.07.016>
- Madasa, A., Orimoloye, I. R., & Ololade, O. O. (2021). Application of geospatial indices for mapping land cover/use change detection in a mining area. *Journal of African Earth Sciences*, 175, 104108. <https://doi.org/10.1016/j.jafrearsci.2021.104108>
- Maldaner, L. F., Corrêdo, L. P., Canata, T. F., & Molin, J. P. (2021b). Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. *Computers and Electronics in Agriculture*, 181, 105945. <https://doi.org/10.1016/j.compag.2020.105945>
- Maldaner, L. F., Molin, J. P., & Spekken, M. (2021a). Methodology to filter out outliers in high spatial density data to improve maps reliability. *Scientia Agricola*, 79, e20200178. <https://doi.org/10.1590/1678-992X-2020-0178>
- Pinty, W., & Verstraete, M. M. (1991). GEMI: A non-linear index to monitor global vegetation from satellites. *Vegetatio*, 101, 15–20. <https://doi.org/10.1007/BF00031911>
- QGIS - Quantum GIS. (2021). *QGIS Geographic Information System*. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>
- Ramos, A. P. M., Osco, L. P., Furuya, D. E. G., Gonçalves, W. N., Santana, D. C., Teodoro, L. P. R., Silva Junior, C. A., Capristo-Silva, G. F., Li, J., Baio, F. H. R., Marcato Junior, J., Teodoro, P. E., & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, 178, 105791. <https://doi.org/10.1016/j.compag.2020.105791>
- Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55, 95–107. [https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7)
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTs. In S. C. Freden, E. P. Mercanti, & M. A. Becker (Eds.), *Third Earth resources technology satellite-1 symposium—Volume I: Technical presentations*, NASA SP-351 (pp. 309–317). NASA.
- Saranya, C. P., & Nagarajan, N. (2020). Efficient agricultural yield prediction using metaheuristic optimized artificial neural network using Hadoop framework. *Soft Computing*, 24, 12659–12669. <https://doi.org/10.1007/s00500-020-04707-z>
- Schwalbert, R. A., Amado, T. J. C., Nieto, L., Varela, S., Corassa, G. M., Horbe, T. A. N., Arroz, C. W., Peralta, N. R., & Ciampitti, I. A. (2018). Forecasting maize yield at field scale based on high-resolution satellite imagery. *Biosystems Engineering*, 171, 179–192. <https://doi.org/10.1016/j.biosystemseng.2018.04.020>

- Silva, D. F. D., Garcia, P. H. D. M., Santos, G. C. L., Farias, I. M. S. C., Pádua, G. V. G., Pereira, P. H. B., Silva, F. E., Batista, R. F., Gonzaga Neto, S., & Cabral, A. M. D. (2021). Morphological characteristics, genetic improvement and planting density of sorghum and corn crops: A review. *Research, Society and Development*, 10, e12310313172. <https://doi.org/10.33448/rsd-v10i3.13172>
- Sripada, R. P., Schmidt, J. P., Dellinger, A. E., & Beegle, D. B. (2008). Evaluating multiple indices from a canopy reflectance sensor to estimate corn N requirements. *Agronomy Journal*, 100, 1553–1561. <https://doi.org/10.2134/agronj2008.0017>
- Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., & Mochizuki, K. (2018). Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. *Journal of Applied Remote Sensing*, 12, 026019. <https://doi.org/10.1117/1.JRS.12.026019>
- Suresh, K., Behera, S. K., Manorama, K., & Mathur, R. K. (2021). Phenological stages and degree days of oil palm crosses grown under irrigation in tropical conditions. *Annals of Applied Biology*, 178, 121–128. <https://doi.org/10.1111/aab.12641>
- Tedesco, D., Moreira, A. B. R., Barbosa Júnior, M. R., Papa, J. P., & Silva, R. P. (2021). Predicting on multi-target regression for the yield of sweet potato by the market class of its roots upon vegetation indices. *Computers and Electronics in Agriculture*, 191, 106544. <https://doi.org/10.1016/j.compag.2021.106544>
- Tedesco, D., Oliveira, M. F., Santos, A. D., Silva, E. H. C., Rolim, G. S., & Silva, R. P. (2021). Use of remote sensing to characterize the phenological development and to predict sweet potato yield in two growing seasons. *European Journal of Agronomy*, 129, 126337. <https://doi.org/10.1016/j.eja.2021.126337>
- Venancio, L. P., Mantovani, E. C., Amaral, C. H., Neale, C. M. U., Gonçalves, I. Z., Filgueiras, R., & Campos, I. (2019). Forecasting corn yield at the farm level in Brazil based on the FAO-66 approach and soil-adjusted vegetation index (SAVI). *Agricultural Water Management*, 225, 105779. <https://doi.org/10.1016/j.agwat.2019.105779>
- Wang, T., Shi, J., Letu, H., Ma, Y., Li, X., & Zheng, Y. (2019). Detection and removal of clouds and associated shadows in satellite imagery based on simulated radiance fields. *Journal of Geophysical Research: Atmospheres*, 124, 7207–7225. <https://doi.org/10.1029/2018JD029960>
- Wu, W. (2014). The generalized difference vegetation index (GDVI) for dryland characterization. *Remote Sensing*, 6, 1211–1233. <https://doi.org/10.3390/rs6021211>
- Xu, Y., Yu, L., Cai, Z., Zhao, J., Peng, D., Li, C., & Gong, P. (2019). Exploring intra-annual variation in cropland classification accuracy using monthly, seasonal, and yearly sample set. *International Journal of Remote Sensing*, 40, 8748–8763. <https://doi.org/10.1080/01431161.2019.1620377>
- Zhangyan, J., Huete, A. R., Kamel, D., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112, 3833–3845. <https://doi.org/10.1016/j.rse.2008.06.006>
- Zhou, X., Kono, Y., Win, A., Matsui, T., & Tanaka, T. S. T. (2021). Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. *Plant Production Science*, 24, 137–151. <https://doi.org/10.1080/1343943X.2020.1819165>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pinto, A. A., Zerbato, C., Rolim, G. de S., Barbosa Júnior, M. R., Silva, L. F. V. da, & Oliveira, R. P. de (2022). Corn grain yield forecasting by satellite remote sensing and machine-learning models. *Agronomy Journal*, 114, 2956–2968. <https://doi.org/10.1002/agj2.21141>