PEANUT LOSS PREDICTIONS USING RANDOM FOREST AND SOIL DATA

Armando L. de Brito Filho[1], Franciele M. Carneiro[2], Gabriel P. Costa[1], Lucas Matheus Agostini[1], Jarlyson Brunno Costa Souza[1], Rouverson P. da Silva[1]

[1]Department of Engineering and Mathematical Sciences, São Paulo State University (Unesp), Jaboticabal, São Paulo, Brazil

[2]Federal Technological University of Paraná (UTFPR), Santa Helena 85892-000, Paraná, Brazil

armando.brito@unesp.br, franmorlin1@gmail.com, pereira.costa@unesp.br, l.agostini@unesp.br, jarlyson.brunno@unesp.br, rouverson.silva@unesp.br

Abstract. *This work aimed to evaluate the Random Forest algorithm performance to predict mechanical digging loss of the peanut crop. Four approaches were tested: only soil texture data, soil texture + TPI, soil texture + TWI, and soil texture + TPI + TWI. The model's performance was evaluated regarding precision (coefficient of determination) and accuracy (mean absolute error). The results found in this work proved promising in predicting peanut digging losses. Our models achieve results with an approximate 100 kg ha-1 prediction error. In addition, incorporating one or more topographical indexes in conjunction with soil texture data as features notably improves the models' precision and accuracy.*

1. Introduction

Currently, in Brazil, the peanut harvesting operation is carried out in a fully mechanized way. However, during this process, significant and unavoidable losses occur. The first stage of mechanized peanut harvesting is known as digging and is responsible for most of the total losses, reaching up to 30% [ZERBATO et al., 2019; SILVA, 2019]. Therefore, the measurement of losses in the mechanized uprooting process is an essential metric since, by obtaining reliable estimates, farmers can adopt adequate management and decision-making strategies for their crops.

Measuring losses facilitates the development of innovative practices and technologies to mitigate them. Thus, several studies have been carried out to evaluate the losses arising from the harvesting processes and propose strategies to minimize them, mainly considering the factors that tend to influence their occurrence. Within this context, soil-related factors, such as physical and textural characteristics, influence the variability of losses [ZERBATO et al., 2017]. Peanut develops its pods in the lower part of the soil, so the plant needs lighter and more aerated soils, which are more easily found in sandy or sandy loam lands. In contrast, clayey soils are heavier, favoring the plant's sound development, but pod losses are more significant at harvest time.

In recent studies, the use of topographic indices, such as the Topographic Position Index (TPI) and the Topographic Moisture Index (TWI), has played an essential role in the development of models for predicting agricultural productivity [Oliveira et al., 2022]. These indices provide valuable information about the physical attributes of the landscape and their influence on agricultural processes. Thus, soil physical factors stand out as an essential variable to explain the variability in the presence of losses. However, to develop accurate predictive models of peanut losses, it is essential to consider not only soil attributes but also the available advanced analytical techniques.

Advances in data processing have enabled the use of complex analytical techniques, such as Machine Learning (ML). Among popular ML methods, Random Forest (RF) is a powerful and versatile algorithm based on independent decision trees, widely used for prediction and classification problems [Jeong et al., 2016]. Approaches using this type of method have shown high potential in agronomic forecasts [Yue et al., 2019, Oliveira et al., 2022]. However, it is unaware of such loss prediction scenarios for the peanut crop.

Integrating advanced data processing techniques plus information on soil characteristics parameters can provide a comprehensive approach to predicting peanut losses. Therefore, the objective of this work was to evaluate the Random Forest algorithm performance to predict mechanical digging loss of the peanut crop.

2. Material and Methods

In this section, we present how the development of the database for training the algorithms will be carried out (Section 2.1), the implementation of the proposed machine learning algorithm (Section 2.2), and finally, the method used to evaluate the performance of the algorithm (Section 2.3). Figure 1 shows a graphic summary of this study.

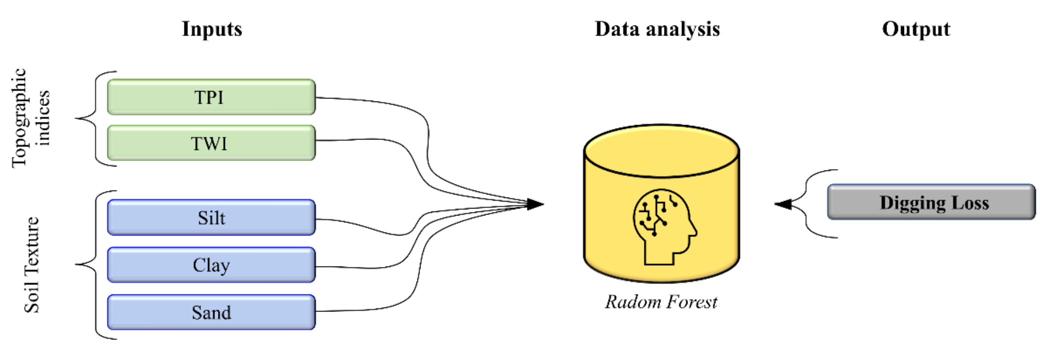

Figure 1. Graphic abstract of the project with the stages of development.

2.1. Study area and database development

The work was carried out in the state of São Paulo - Brazil, in three agricultural production areas (Figure 2), for commercial purposes, in the 2022/2023 harvest. Two cultivars with different cycles and two soil conditions were harvested to express the existing variability in the field better, as shown in Table 1.

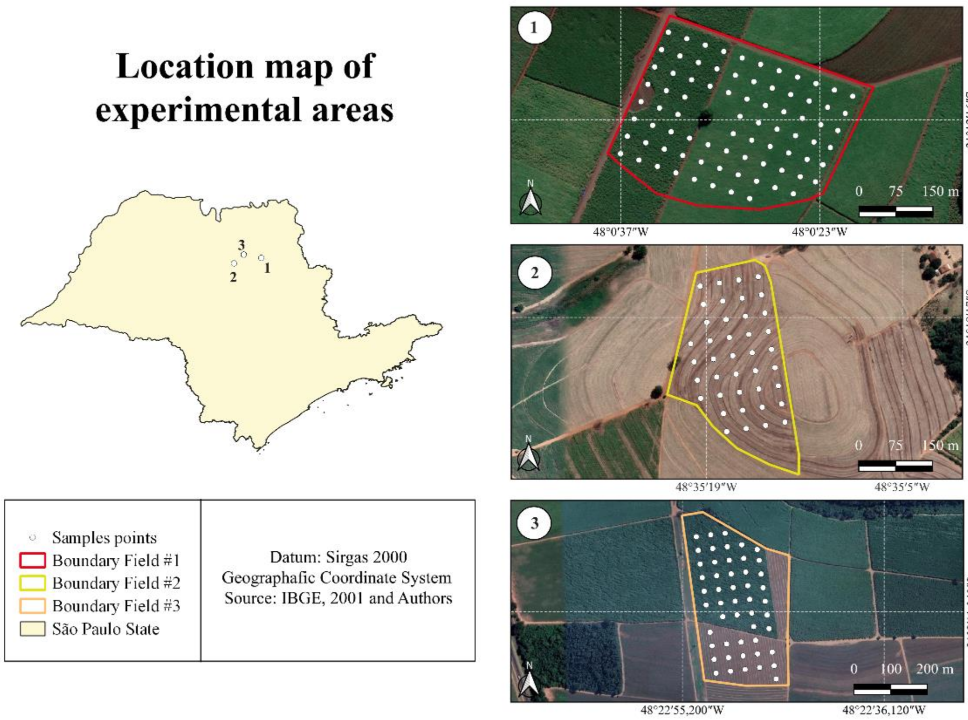

Figure 2. Location map of the experimental area in a peanut agricultural.

Table 1. Description and characteristics of the study areas

Field	City	Peanut cultivars	Soil texture type	Sowing
1	Sertãozinho	OL3	Clay soil	Early sowing
2	Taquaritinga	IAC503	Sandy soil	Late sowing
3	Taiuva	OL3	Sandy soil	Early sowing

The loss data was sampled using a regular grid of 80 experimental plots, spaced 40 x 40 m apart, in field # 1, 70 experimental plots, spaced 35 x 35 m apart, in field # 2, and 50 experimental plots, spaced 40 x 40 m apart, in field # 2. Within each plot, soil texture information (sand, silt, and clay) was collected referring to the 0,00 – 0,20 m layer, representing the average for the root development of the peanut crop.

The loss evaluations were carried out using rectangular frames with an area of 2 m2. The quantifications of total losses were carried out with the sum of visible and invisible losses. The collection was carried out by removing all the pods on the soil surface after the formation of the peanut windrows, which then represented the visible losses in the uprooting. In order to collect the invisible losses during uprooting, it was necessary to excavate the soil up to 15 to 20 cm, the layer in which the peanut pods are normally located.

The Topographic Position Index (TPI) and Topographic Wetness Index (TWI) were generated by the QGIS® software, version 3.28.9. In order to create the indices, it is necessary to have information from the Digital Elevation Model (DEM), which was obtained through the "Open Topographic DEM" plugin, in which the DEM with a spatial resolution of 30 m from the Copernicus Global DSM platform was chosen. After that, with the DEM, the TPI was created using the "Multi-Scale Topographic Position Index" plugin from the SAGA platform. As for the creation of the TWI, it was necessary to have, in addition to the DEM, slope, and catchment area, both parameters created in QGIS®, and finally, with the DEM, slope, and catchment area parameters, the TWI was created using the "Topographic Wetness" plugin Index" of the SAGA platform.

2.2. Implementation of the Random Forest algorithm

The RF algorithm was processed using the Python, 3.8.10 version, programming language in the Jupyter Notebook environment. The best combination of the number of trees (n_estimators) and maximum tree depth (max_depth) was selected using the GridSearch tool. At this processing stage, our dataset will be divided into 80% for training and 20% for validation. For model development, we worked with four approaches: soil texture data only, soil texture + TPI, soil texture + TWI, and soil texture + TPI + TWI. In addition, a test stage was carried out in an area different from the training area.

2.3. Random Forest algorithm performance

The RF algorithm performance was evaluated through precision, by the Coefficient of Determination (R^2), and accuracy, by the Mean Absolute Error (MAE).

3. Result and discussion

Table 2 shows the performance results of the loss prediction model with the RF algorithm. The performance comes from separating 20% of the original data set used for training. The model that showed greater precision and accuracy in its predictions was trained with soil texture data and the topographic indices TPI and TWI, reaching values of $R^2 = 0.78$ and MAE = 93.05.

The approach that presented the lowest performance compared to the other combinations was the one that only relied on soil texture data. However, the model achieved precision with $R^2 = 0.68$ and MAE = 119.99 kg ha^{-1} accuracy. Soil texture combinations with TWI or TPI indices obtained similar results.

Table 2. Performance of valid step with Random Forest model

Approaches	Metrics	
	R^2	MAE (kg ha^{-1})
Texture	0.65	119.99
Texture + TPI	0.75	99.22
Texture + TWI	0.75	100.69
Texture + TPI + TWI	0.78	93.05

Soil texture data were sufficient to generate a loss prediction model. However, adding information from topographic indices contributed to developing a more precise and accurate model. After the inclusion of these variables, the model had an increase of 0.1 and 0.13 in precision and a decrease in the error of approximately 20 and 26 kg ha^{-1}, increasing accuracy.

Soil characteristics parameters are essential indicators for measuring losses. Soils with different textures present different amounts of losses [ZERBATO et al., 2017]. In addition, topographic indices provide essential information about the characteristics of the earth's surface. The TPI provides information on the spatial distribution of landforms, which may indicate points with more excellent slopes or slopes within agricultural areas. The presence of such surfaces tends to influence the performance of the machinery in the field, causing losses.

On the other hand, TWI is an index correlated with soil moisture, reflecting water availability for crop growth and development [Yang et al., 2021]. Soil moisture is an essential indicator of the quality of the harvesting operation since it can influence the appearance of losses [Behera et al., 2008; Zerbato et al., 2014]. In addition, according to the authors Ince and Guzel (2003), soil moisture has an exponential relationship with the force of peanut detachment, so when the water content decreases, this force decreases; consequently, the total losses in the harvest increase.

In Figure 3, we observe the performance of the models created in a test area in conditions different from those used in training. In general, the models showed the same behavior as the test stage. In which the best results were from the integration of texture data with data from topographic indices.

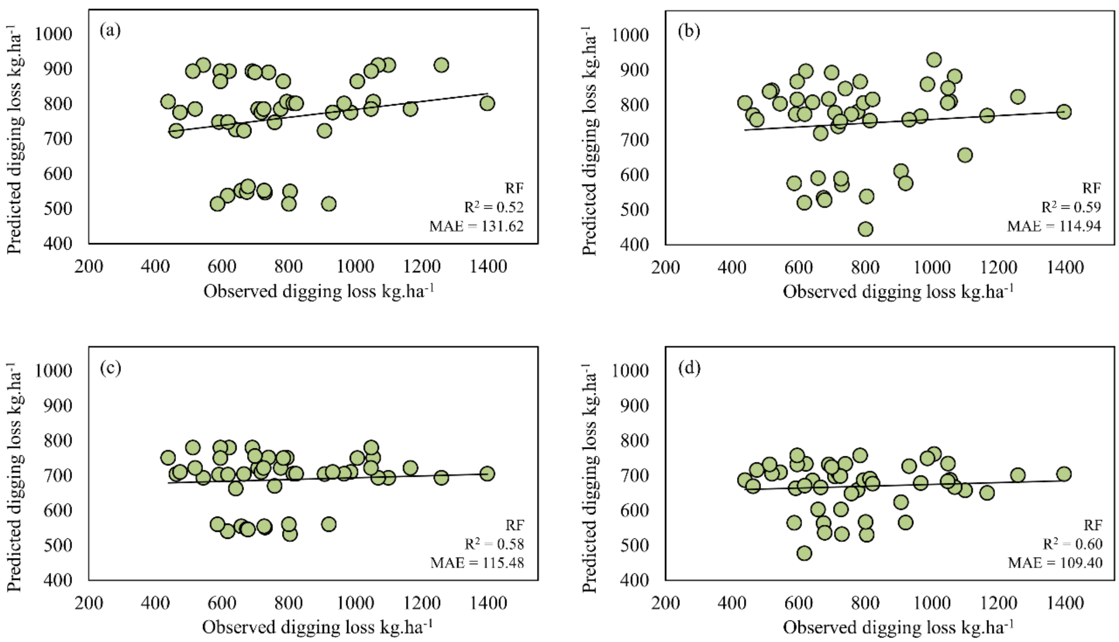

Figure 3. Graphics of observed versus predicted digging loss values using different approaches. (a) just texture; (b) texture + TPI; (c) texture + TWI; (d) texture + TPI + TWI.

The model generated only with soil texture data showed the lowest performance, reaching $R^2 = 0.52$ and MAE = 131.62 kg ha^{-1} (Figure 3 a). However, it should be noted that mechanized peanut harvesting operations present high variability in their loss data [Santos et al., 2019; Zerbato et al., 2019], which may have influenced the performance of the models.

The results found in this work proved promising in predicting peanut digging losses. The current approaches to measuring losses demand time and significant efforts to collect information. Our models achieve results with an approximate 100 kg ha^{-1} prediction error. In addition, with the prediction of losses in the mechanized peanut harvesting process, farmers can adopt appropriate management, decision-making, and resource allocation strategies.

4. Conclusion

The Random Forest (RF) algorithm showed excellent performance for predicting peanut digging losses. Moreover, approaches incorporating one or more topographical indexes in conjunction with soil texture data as features notably improves the models' precision and accuracy.

The results found in these researches can be considered promising in indirectly predicting losses in the peanut harvest. Based on this, future research can expand the amount of input data into the model, with the addition of information regarding the agronomic characteristics of the crop since these also tend to influence the amount of losses.

5. References

Behera, B. K., Behera, D., Mohapatra, A. K., Swain, S., Goel, A. K. (2008) Performance evaluation of a Bullock drawn groundnut digger. Environment and Ecology, v.26, p.1226-1229.

Ince, A. and Guzel E. (2003). Effects of gynophore breaking resistance on losses in mechanized peanut harvesting. In: International conference on crop harvesting and processing, Louisville, Kentucky. ProceedingsÉ St Joseph: ASABE. 2003. p. 1103.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Kim, S. H. (2016). Random forests for global and regional crop yield predictions. PloS one, 11(6), e0156571

Oliveira, R. P., Barbosa Júnior, M. R., Pinto, A. A., Oliveira, J. L. P., Zerbato, C., Furlani, C. E. A. (2022). Predicting sugarcane biometric parameters by UAV multispectral images and machine learning. Agronomy, 12(9), 1992.

Santos, A. F., Silva, R. P., Zerbato, C., Menezes, P. C., Kazama, E. H., Paixão, C. S. S., Voltarelli, M. A. (2019). Use of real- time extend GNSS for planting and inverting peanuts. Precis. Agric. 2019, 20, 840–856.

Silva, R. P. (2019). Colheita mecanizada de amendoim. In: Silva R. P.; Santos A. F.; Carrega W. C. (Eds.) Avanços na produção de amendoim. Jaboticabal: FUNEP, p. 129–141.

Yang, M., Wang, G., Lazin, R., Shen, X., Anagnostou, E. (2021) Impact of planting time soil moisture on cereal crop yield in the Upper Blue Nile Basin: A novel insight towards agricultural water management. Agricultural Water Management, v. 243, n. February 2020, p. 106430.

Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. ISPRS Journal of Photogrammetry and Remote Sensing 150:226–244.

Zerbato, C., Furlani, C. E. A., Silva, R. P., Voltarelli, M. A., Santos, A. F. (2017). Statistical control of processes aplied for peanut mechanical digging in soil te xtural classes. Engenharia Agrícola, 37(2), 315-322.

Zerbato, C., Furlani, C. E. A., Oliveira M. F. D., Voltarelli, M. A., Tavares, T. O., Carneiro, F. M. (2019). Quality of mechanical peanut sowing and digging using autopilot. Revista Brasileira de Engenharia Agrícola e Ambiental, v. 23, n. 8, p. 630-637.

Zerbato, C., Silva, V. F., Torres, L. S., Silva, R. P. D., Furlani, C. E. (2014). Peanut mechanized digging regarding to plant population and soil water level. Revista Brasileira de Engenharia Agrícola e Ambiental, 18, 459-465.