ICPA
17th INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE
ConBAP 2026
CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO E DIGITAL
13-16 July 2026 | Porto Alegre, Brazil

ESTIMATING PEANUT LOSSES USING MACHINE LEARNING WITH SOIL AND WEATHER DATA

***Armando Lopes de Brito Filho*[1*], *Franciele Morlin Carneiro*[2], *Gabriel Pereira Costa*[3], *Breno dos Santos Silva*[3], *Pedro Henrique Nogueira Gusmão*[3], *Rouverson Pereira da Silva*[3]**

[1]Western Paraná State University (UNIOESTE), Cascavel *campus*, Paraná State, Brazil.
[2]Federal Technological University of Paraná (UTFPR), Santa Helena *campus*, Paraná State, Brazil.
[3]São Paulo State University (Unesp), Jaboticabal *campus*, São Paulo State, Brazil.

*email:armandolopesbrito@gmail.com

Abstract.

Mechanized peanut harvesting is an important phase of the production system, directly affecting both production costs and crop yield. However, due to interactions among soil conditions, plant characteristics, and machine performance, this operation is carried out under challenging conditions that may result in high levels of loss. These losses are classified as visible when pods remain on the soil surface after digging and as invisible when they are incorporated into the soil profile, making them more difficult to detect and quantify. The high spatial variability of these losses highlights the need for more robust and precise estimation methods to support management strategies aimed at continuously reducing operational and economic losses. This study aimed to evaluate the performance of machine learning algorithms in estimating peanut losses associated with mechanical peanut digging. Five machine learning (ML) algorithms were assessed: Multiple Linear Regression (MLR), K-Nearest Neighbors (KNN), Support Vector Regression (SVR), eXtreme Gradient Boosting (XGB), and Random Forest (RF). These algorithms were used to estimate peanut losses based on soil and weather data collected from eight commercial fields in São Paulo State, Brazil. The models' performance was evaluated using mean absolute error (MAE) and the coefficient of determination (R^2). The results revealed high spatial heterogeneity of losses, confirming the complex and dynamic nature of the harvesting process under field conditions. Overall, the models exhibited similar MAE values, indicating controlled average errors and an ability to estimate the general magnitude of losses. However, R^2 values were low, particularly for invisible losses, indicating limited capacity to explain the spatial variability observed in the dataset. This discrepancy indicates that while the algorithms can capture average trends, they struggle to represent more complex spatial patterns, especially those related to subsurface losses. In conclusion, machine learning algorithms show potential as supportive tools for estimating losses in mechanized peanut harvesting, particularly for overall operational monitoring. However, their capacity for spatial generalization remains limited, especially regarding invisible losses. Expanding and diversifying the training database is a great strategy to improve the models' estimated performance. Therefore, this study contributes to the advancement of digital agriculture in mechanization, highlighting the potential and current limitations of artificial intelligence-based approaches in peanut harvest management.

Keywords. *Agricultural mechanization, Artificial intelligence, Digital agriculture, Harvest.*

Introduction

Mechanized peanut harvesting is one of the most critical stages of production, as it is directly associated with yield losses and operational efficiency of the crop (SILVA, 2019; ZERBATO et al., 2019). Models such as Random Forest Regressor (RF), XGBoost (Extreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), CatBoost (CatBoost Regressor), KNN (K-Neighbors Regressor), SVR (Support Vector Regression), and Gradient Boosting Regressor (GBR), trained with data from commercial peanut cultivation areas in the state of São Paulo, were evaluated. The results showed high spatial variability in losses, a characteristic frequently observed in the mechanized harvesting of this crop (ORMOND et al., 2018; ZERBATO et al., 2017), and demonstrated that, although the models present controlled mean errors, their ability to represent spatial variability, especially invisible losses, is still limited. However, the algorithms showed potential to estimate the average magnitude of losses, which could support decision-making, operational monitoring, and mechanized harvesting planning. In this context, the present study aimed to evaluate the performance of different machine learning algorithms in estimating visible and invisible losses associated with mechanized peanut harvesting, using edaphic, topographic, climatic, and agronomic variables.

Material and Methods

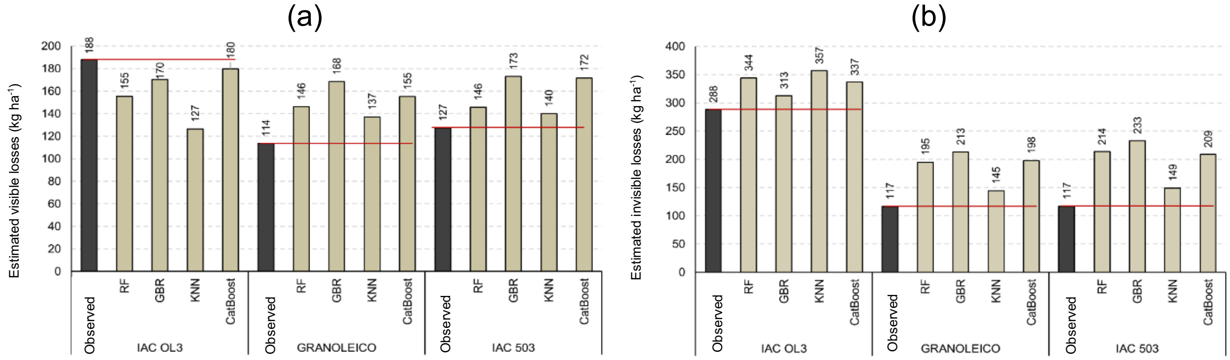
This research used data on losses from mechanized peanut harvesting collected in commercial areas in the Jaboticabal region, SP, over five agricultural seasons. Visible and invisible losses were quantified after the mechanized harvesting operation, using 2 m² sampling frames. In addition to loss data, soil information, topographic indices, and climatic variables were incorporated. The topographic indices TPI (Topographic Position Index) and TWI (Topographic Wetness Index) were generated in QGIS software from digital elevation models, while climatic data were obtained from the NASA POWER platform and subsequently integrated into the database for analysis. After consolidating the database, the following machine learning algorithms were evaluated: RF, GBR, KNN, LightGBM, CatBoost, XGBoost, and SVR. The models were trained using four input approaches that combined information on soil, topography, climate, and crop characteristics. The dataset was divided into 80% for training and 20% for validation, with the performance of the models evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE), aiming to analyze their predictive and generalization capacity in estimating visible and invisible losses from mechanized peanut harvesting.

Results and Discussion

The results showed high spatial variability in visible and invisible losses during mechanized peanut harvesting, characterized by wide data dispersion and frequent outliers. This variability was influenced by soil, relief, climate, and crop conditions, reinforcing the complexity of the mechanized harvesting process (ORMOND et al., 2018; ZERBATO et al., 2017). Invisible losses showed even more heterogeneous behavior, highlighting the influence of interactions between plant, soil, and machine. These results demonstrate that conventional approaches have limitations in capturing loss dynamics, underscoring the need for more robust analytical tools to support field monitoring and decision-making.

In general, the machine learning algorithms showed similar performance in terms of prediction errors (MAE and RMSE), but with low R^2 values, especially for invisible losses, indicating difficulty in representing the observed spatial variability (Fig 1). Although the models demonstrated limited generalization to new areas and cultivars, they were able to adequately estimate the average magnitude of losses, maintaining errors within the range of the natural variability reported in the literature (SANTOS et al., 2019; SANTOS et al., 2021; ZERBATO et al., 2019). The results indicate that expanding and diversifying the database, and incorporating new variables that capture operational and environmental conditions, can improve the predictive capacity and practical application of these models for the planning and management of mechanized peanut harvesting.

Fig 1. Average of visible (A) and invisible (B) losses observed and estimated by machine learning models in different peanut cultivars.



Conclusion

It was concluded that losses in mechanized peanut harvesting exhibit high spatial variability, reflecting the complex interactions among soil, plants, machines, and the environment. The study demonstrated that machine learning algorithms have limited capacity to represent this variability, especially for invisible losses, although they show good accuracy in estimating the average magnitude of losses. The results indicate that expanding the database and incorporating more representative variables can improve the generalization capacity of the models, thereby enhancing their potential as decision-making, monitoring, and risk-management tools in mechanized peanut harvesting.

References

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. *Precision Agriculture*, *20*, 840–56.

Santos, A. F., de Oliveira, L. P., de Oliveira, B. R., Ormond, A. T. S., & da Silva, R. P. (2021). Can digger blades wear affect the quality of peanut digging? *Revista Engenharia na Agricultura, 29*, 49-57.

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

Ormond, A. T. S., Santos, A.F., Alcantara, A. S., Zerbato, C., & Furlani, C. E. A. (2018). Tillage Interference in the Quality of Peanut Mechanized Harvest. *Engenharia Agrícola*, *38*, 251–59.

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 Textural Classes. *Engenharia Agrícola*, *37*, 315–22.

Zerbato, C., Furlani, C. E. A., Oliveira M. F. de, 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*, *23*, 630-637.