EasyChair Preprint

№ 13705

Predicting Orange Harvest Using Image Analysis and RandomForestRegressor

Alessandro Akira Saito and Milton Faria Jr

June 19, 2024

Predicting Orange Harvest Using Image Analysis and RandomForestRegressor

Alessandro Akira A. Saito, Milton Faria Jr.
Computer Engineering
Universidade de Ribeirão Preto
Ribeirão Preto, SP, Brazil
Correspondence: mfaria@unaerp.br

Abstract*— AI in agriculture offers numerous benefits, from increased productivity and efficiency to waste reduction and sustainable practices. This revolutionary technology processes large volumes of data, providing valuable insights on climate, plant health, water management, and logistics. AI also enables the automation of agricultural tasks, such as harvesting and pesticide spraying, making the process more efficient and cost-effective (RENDA, 2023, SCHMIDHUBER, 2015). The "Prediction of Orange Harvest Based on Flowering Image Analysis" is a fascinating and innovative topic that combines agriculture with modern technology. This study aims to develop efficient methods to predict orange production using images of tree flowering. Flowering is a crucial indicator of the productive potential of an orange tree, yet the relationship between flowering and final fruit production is complex and influenced by many factors, including climatic conditions, tree health, and agricultural practices. (REZENDE, 2004 , ELISANDRA 2005, WILLIAMS, 1969)

Keywords—deep learning, RandonForest, predicting orange

I. Introduction

Brazil's economy is heavily influenced by agriculture, which accounts for approximately 25% of the national GDP and over 40% of exports, according to the Center for Advanced Studies in Applied Economics (CEPEA, 2023). The agricultural sector generates income, employment, and foreign exchange, contributing to food security and environmental preservation. However, Brazilian agriculture faces numerous challenges, such as climate variability, international competition, technological innovation, and social inclusion. (DECICINO 2023, JARDIMANIMAL, 2023)

Agriculture is one of the oldest and most fundamental industries, providing food, fibers, and raw materials essential for human survival and well-being. It faces significant challenges as the global population grows, the climate becomes more unpredictable, and natural resources become scarcer. In this context, Artificial Intelligence (AI) has emerged as a transformative tool, playing a crucial role in modernizing and optimizing agriculture. (GOODFELLOW 2016)

The primary objective of this study is to develop an algorithm capable of analyzing photos of orange trees and weather data to predict harvest outcomes at the end of the season. Brazil is the world leader in orange production, holding about 73% of the global orange juice market. The main importers of Brazilian orange juice are the European Union and the United States. While citrus production is concentrated in São Paulo and the Triângulo Sul/Mineiro region, the Northeast region, specifically Sergipe and Bahia, has significant participation, with over 80% of the cultivation area.

II. Methods

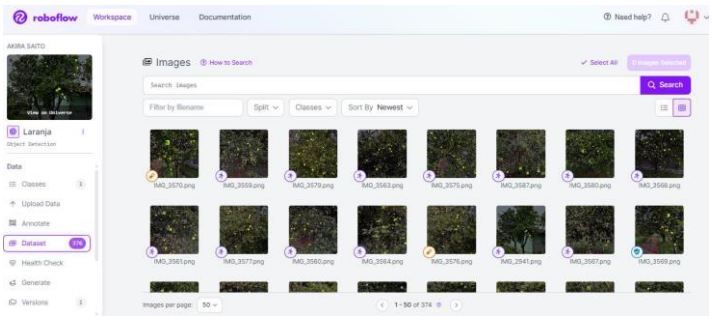
Data Collection: We used a dataset of 376 images of orange trees, captured using a smartphone with a 48 MP wide-angle lens. These images were stored on the Roboflow platform, which facilitates the development of computer vision applications. The dataset was divided into training (70%), validation (20%), and test (10%) sets to ensure effective model training and evaluation (Fig. 1).

Detailed climatic information, including average temperature and precipitation, was obtained from the INMET (National Institute of Meteorology) website. Additionally, data on the daily amount of sunlight, which is crucial for orange tree growth, was factored in by considering the latitude of the study location. This comprehensive climatic data is fundamental for developing accurate regression models that predict orange production by incorporating environmental variables that directly affect plant growth and productivity.

The YOLOv8 deep learning algorithm was employed to analyze the images. YOLOv8 (You Only Look Once version 8) is known for its efficiency in object detection tasks. Each image in the dataset was accompanied by a text file containing bounding box information that specified the coordinates of the objects of interest, which in this case were the orange blossoms. This format is essential for YOLOv8 to interpret and process the images accurately during training.

The RandomForestRegressor algorithm, an advanced AI method, was applied to predict orange harvest based on the analyzed images and climatic data. Random forests are ensemble learning methods that operate by constructing multiple decision trees during training and outputting the mean prediction of the individual trees, thus improving predictive accuracy and controlling overfitting. The model was evaluated using R-squared (R^2) and Mean Squared Error (MSE) metrics, achieving an R^2 of 0.9358 and an MSE of 16.56.

Fig. 1. *Database hosted on Roboflow.*



The RandomForestRegressor model predicted that the 2024 harvest in northern São Paulo would yield approximately 125,111 boxes of oranges. In comparison, the Fundecitrus report released on May 23, 2024, estimated a harvest of approximately 90,390 boxes, indicating a discrepancy of 34,721 boxes between the two models.

Fig. 2. Table.*by years*

Year	Harvest	Trees	Temp(°C)	Precip(L)
2015	90570	37963	22,03	0,15
2016	56680	39604	21,70	0,14
2017	124550	39290	23,69	0,16
2018	65920	39324	22,25	0,13
2019	119230	40297	24,31	0,15
2020	70230	39726	22,99	0,13
2021	84880	39665	21,75	0,15
2022	89740	39183	21,28	0,21
2023	110930	38932	22,20	0,22

III. CONCLUSION

Brazil's economy is heavily influenced by agriculture The results underscore the importance of integrating AI with traditional agricultural practices. The high R^2 value indicates the model's robustness in explaining the variance in orange production based on the input variables. The discrepancy with the Fundecitrus report highlights the need for continuous model refinement and validation against real-world data. Factors such as image quality, variability in flower appearance, and complex environmental influences can affect prediction accuracy. Future research could focus on enhancing image analysis techniques and incorporating additional climatic and biological variables to improve model performance.

The study highlights the significant impact of climate on orange production, with conditions like water stress and high temperatures negatively affecting yield. The developed AI-based system serves as a valuable tool for farmers, providing crucial information that can aid in decision-making. By anticipating adverse climatic conditions, farmers can adopt preventive measures and management strategies to minimize losses and optimize production. The non-destructive harvesting method integrated into the project also reduces waste, promoting a more sustainable and efficient approach to orange harvesting.

The model was evaluated using R-squared (R^2) and Mean Squared Error (MSE) metrics, achieving an R^2 of 0.9358 and an MSE of 16.56.

REFERENCES

[1] CEPEA, PIB do Agronegócio Brasileiro - Centro de Estudos Avançados em Economia Aplicada - CEPEA-Esalq/USP.. Disponível em: <https://www.cepea.esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx>. Acesso em: 25 out. 2023.

[2] REZENDE, T. H.; BARBOSA, J. C. Estimativa do número de frutos por amostragem de parte da copa em laranjeiras: [s.l.] Revista Brasileira de Fruticultura, 2004.

[3] DECICINO, R. Fruticultura no Brasil: Importância econômica para o país. Disponível em: <https://educacao.uol.com.br/disciplinas/geografia/fruticultura-no-brasil-importancia-economica-para-o-pais.htm>. Acesso em: 27 nov. 2023

[4] JARDIMANIMAL. Cultivo de laranja, saiba tudo do plantio ate a colheita. 2019. Disponível em: <https://jardimanimal.com.br/cultivo-de-laranja-saiba-tudo-do-plantio-ate-a-colheita/>. Acesso em: 27 nov. 2023.

[5] ELISANDRA, S. et al. Modelos agrometeorológicos para estimação do número de frutos por planta em laranjeira 'Pêra' na região de Limeira-SP, Brasil. v. 13, n. 3, p. 412–422, 1 jan. 2005.

[6] WILLIAMS, S. R. Forecasting Florida Citrus Production, Methodology and Development. Florida: Florida Crop and Livestock Reporting Service, 1969. 54p.

[7] JESSEN R. J. Determining the Fruit Count on a Tree by Randomized Branch Sampling. Iwoa: Iowa State College, 1955. 11p.

[8] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. Cambridge, Massachusetts: The Mit Press, 2016.

[9] SCHMIDHUBER, J. Deep learning in neural networks: An overview. ScienceDirect, v. 61, p. 85–117, jan. 2015.