2022-06

# Sunflower yield prediction based on high resolution satellite imagery

Pejak Branislav, Marko Oskar, Lončar-Turukalo Tatjana, Ljubičić Nataša, Crnojević Vladimir

# SUNFLOWER YIELD PREDICTION BASED ON HIGH RESOLUTION SATELLITE IMAGERY

Branislav Pejak[1*], Oskar Marko[1], Tatjana Lončar-Turukalo[2], Nataša Ljubičić[1], Vladimir Crnojević[1]

[1]*BioSense Institute, University of Novi Sad, Novi Sad, Serbia*
[2]*Faculty of Technical Science, University of Novi Sad, Novi Sad, Serbia*

Corresponding author: *branislav.pejak@biosense.rs

**Abstract**

High-tech agricultural services and precision agriculture based on remote sensing are methods through which farmers minimize the use of inputs such as water and fertilizer and at the same time enhance productivity, quality and yield. Big data collected by satellite platforms is facilitated for real-time decision making, which is one of the key objectives of modern agriculture. Yield prediction derived from satellite imagery can be used to increase the efficiency of agricultural production, selective harvesting, logistics, price forecasting, storage planning, import/export strategies, etc.. The objective of this study was to evaluate crop regression models at different points in time using multispectral satellite images. The imagery used in this study was acquired by ESA's Sentinel-2 satellite which provides high-resolution (10 m x 10 m per pixel) satellite images every 5 days. A total of 7 cloud-free Sentinel-2 images were obtained between 04/06/2017 and 03/08/2017. Each of those images contains 13 multispectral channels in the visible, infrared and short-wave infrared part of electromagnetic spectrum. For model training we used both raw data and vegetation indices such as Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Soil-Adjusted Vegetation Index (SAVI) derived from it. In order to adjust the data to the regression model, it was necessary to calculate the average value of both raw data and vegetation indices for each parcel. The ground-truth data and information about yields were collected for the 2017 season on a farm located in the Vojvodina region in northern Serbia. The database consisted of information about the location and sunflower yield for 31 fields which cover 715 ha of arable land. Very dry season caused a huge variation of sunflower yield in the range of 1.6 t/ha to 4 t/ha and the difference was greatest at parcels with irrigation systems. Sunflower yields were predicted using different models such as Random Forest, Support Vector Machine and Artificial Neural Network. Their performance was evaluated using leave-one-out cross-validation strategy. The best result was achieved using Support Vector Machine with polynomial kernel function where the mean square error was 0.27 t/ha and the highest Pearson correlation coefficient was 0.76.

**Keywords:** Yield prediction, Machine learning, Sunflower, Vegetation indices, Sentinel-2